

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

GERM-CELL TUMOUR GENOMICS

The evolution of chemoresistance
in germ-cell cancers PAGE 114

PLANETARY SCIENCE

PLUTO'S ICY HEARTLAND

A closer look at the
Sputnik Planitia basin

PAGES 42 & 86

CELL BIOLOGY

INSIDE MITOCHONDRIA

The dynamics of the energy-
producing organelle

PAGES 43, 74 & 139

GLOBAL FOOD SUPPLY

MAKING MORE OF FOOD

Focused nutrition research
and reliable metrics are key

PAGES 30 & 33

NATURE.COM/NATURE

1 December 2016 £10

Vol. 540, No. 7631



9 770028 083095

THIS WEEK

EDITORIALS

FOOD Chewing on the French distaste for doggy bags **p.8**

WORLD VIEW Post-truth: a guide for the perplexed **p.9**



HEALTH STATUS Social standing influences immune system **p.10**

Break out of the echo chamber

It is crucial to fight discrimination in all its forms, but academia does itself a disservice by excluding conservatives from discourse on societal change.

According to surveys and statistics, most *Nature* readers place themselves on the liberal left of the political spectrum. So two items published in *The New York Times* in the past few weeks will have provoked both consternation and debate. Many of our readers, for example, must have shaken their heads in despair while reading the transcript of an interview with US president-elect Donald Trump last week. The other item, an opinion piece entitled ‘The end of identity liberalism’ probably had many academics simply shaking.

The article by Mark Lilla, a researcher at Columbia University in New York City who specializes in the history of Western intellectual, political and religious thought, called for an end to what he described as an overemphasis by liberals on racial, gender and sexual identity politics. He believes that this focus distracts from core fundamental concepts of democracy and so weakens social cohesion and civic responsibility.

In short, he asserted that many progressives live in bubbles; that they are educationally programmed to be attuned to diversity issues, yet have “shockingly little to say” about political and democratic fundamentals such as class, economics, war and policy issues affecting the common good. Of direct relevance to the US election, he argued that the excessive focus on identity politics by urban and academic elites has left many white, religious and rural groups feeling alienated, threatened and ignored in an unwelcoming environment where the issues that matter to them are given little or no attention.

What is urgently needed, said Lilla, is for US liberalism to refocus on educating all citizens on broader issues that unite people, and on core values of democracy, governance and the major forces shaping international politics — and for the liberal press to educate itself about neglected parts of the country and what matters to people living there.

Lilla argues, perhaps unconvincingly, that fixating on the concerns of particular groups has been divisive, and he calls instead for a focus on unifying issues that affect the majority of people in the United States, with highly charged narrower issues such as sexuality and race tackled with a more-measured sense of scale. But it need not be a trade-off.

The article comes at a time when many in science and academia are rightly worried that Trump’s odious racist, sexist and anti-intellectual remarks during his campaign risk unacceptably broadening the limits of acceptable discourse — and freeing and normalizing people’s worst base instincts and a rhetoric of hate. Not surprisingly, the column has been controversial and has sparked vigorous debate.

But the discussion echoes points made earlier this year by *New York Times* columnist Nicholas Kristof, directed at academics. Kristof, who has long championed diversity issues and so can hardly be accused of conservative bias, argued in a column entitled ‘A confession of liberal intolerance’ that academics are often selectively tolerant, but are intolerant when it comes to considering conservative or religious viewpoints.

Nature has long championed the rights and representation of women and minority groups in science — and will continue to do so. We have also made efforts to reach out and offer a platform to researchers with

diverse political views. We will continue to do that too. And we will engage with and report on efforts by social scientists and others to untangle and properly understand contemporary populist movements. There is already a considerable body of literature on these phenomena and social scientists must weigh in more heavily to inform public debate and vigorously challenge misconceptions — on all sides.

Both articles, although perhaps overstating the case, offer food for thought. They highlight that confirmation bias is rife in all walks of life, including the practice of research and the political viewpoints of academic liberals. No one should kid themselves that they are immune.

Kristof also argued that the low and plunging representation of conservatives and evangelicals on US faculties, and bias against these groups, is itself impoverishing intellectual diversity and discourse. He pointed to an effort to change this state of affairs: the Heterodox Academy, a website set up by centrist social psychologist Jonathan Haidt of New York University to advocate tangible remedies. His column did not go down well with liberals. “You don’t diversify with idiots,” stated one of the most highly recommended comments.

Academics must be vigilant and resist normalization of Trump’s crude vision of society, but must also look in the mirror. A significant chunk of the US population voted for Trump. Are some bigots and racists? Yes; but most aren’t, and progressive academic liberals can’t simply dismiss them as retrograde. More unites Americans than divides them, and building on that common ground is the best antidote to extremism. ■

“Confirmation bias is rife in all walks of life, including the practice of research.”

Critics’ choice

Post-publication scrutiny of papers is essential for science — but it should be done politely.

To err is human. But to catch that error — does that take a computer? That’s a question that psychologists have been wrestling with in recent months, as automated software has been checking their published findings on a huge scale.

The automated review of basic features of scientific papers marks a new front in the battle for research reproducibility, and one that has split the community. The divide must be bridged before it becomes too wide, and that will require criticism to be both offered and received in the true spirit of academic enquiry.

As we report in a Toolbox on page 151, psychologists discovered starting in August that someone — or something — was commenting

on the quality of thousands of their published papers. The comments were left on PubPeer — a website for post-publication review that often hosts anonymous allegations of image manipulation. These can lead to retractions and even, according to at least one lawsuit, to an exciting job offer being rescinded.

In the case of the large-scale comments, the posts had been generated by an algorithm that pointed to potential errors in reported *P* values, measures of statistical significance that are too often used to decide whether results are worth publishing. The program, called statcheck, posted analyses of more than 50,000 papers. It sometimes erroneously tagged correct results as potential errors, and it identified many errors that were real but trivial. It also found instances in which *P* values that had been reported as reaching a threshold for statistical significance were actually just shy of it. Although a few authors have posted explanations and corrected results on PubPeer, none of the posts have to *Nature's* knowledge resulted in any formal corrections or retractions.

Some researchers were confused and upset by the mass fact-check; leaders in the psychological community warned that such projects unduly threaten the reputations of individual researchers and even the field. A former head of the Association for Psychological Science in Washington DC wrote a column decrying the use of “uncurated” social media for personal attacks and harassment. A controversial early draft accused research critics of “methodological terrorism”; it was later revised. Another group of researchers launched a petition that called for discussions to stay polite, but also argued that “the freedom to express legitimate criticism must take priority and be protected”.

To be sure, the automated statcheck comments were lacking some useful context, and the algorithm is far from perfect. But much of the negative reaction has less to do with the ins and outs of a simple computer program than with the importance that people place on scientific papers. These are the currency of funding, tenure and prestige, so any challenges come across as threats to careers and reputation.

The implicit assumption that academic papers must adhere to an impossible standard of perfection does science a horrible disservice.

As *Nature* has pointed out before, the scientific paper is a marker on the way to scientific progress, not itself a destination. Scrutiny of papers is therefore to be welcomed, if only to check that the signposts are pointing in the correct direction. New knowledge arrives constantly to correct and displace the old. It is a messy process, full of acrimonious discussions and painful realizations, but necessary. Errors must be rooted out.

The appropriate reaction depends on the nature of the error. Insightful reasoning can lead to incorrect conclusions that still advance science. A 1996 study of a meteorite that had landed in Allan Hills,

“Anyone who finds flaws should seek corrections with diplomacy and humility.”

Antarctica, argued that elongated nanometre-scale blobs in the rock were the fossils of alien bacteria. Subsequent abiotic explanations felled each argument in turn. But the study breathed life into the field of astrobiology.

Carelessness and avoidable errors will not have such positive effects. Revelations of typos and biased reasoning should make authors uncomfortable. Before submitting their work, they should take on the responsibility of reexamining manuscripts for simple details and limits to their conclusions, and should invite colleagues to do the same. (Peer review improves the scientific literature both by giving papers more credibility and by forcing authors to do just this.)

Even so, errors will make their way into the literature. Anyone who finds flaws should seek corrections with diplomacy and humility. A gloating sense of ‘gotcha’ does not help to provide constructive criticism; some ill-considered phrases have caused lasting damage. But many scientists use their blogs for credible, restrained, nuanced criticism, often engaging the authors whose works are criticized.

Sharing and discussion of scientific work has changed drastically in a world of blogs, online repositories and Twitter. The fact remains, however, that self-correction is at the heart of science. Critics — curated or not — should be courteous, but criticism itself must be embraced. ■

Serving suggestion

Reducing food waste requires a change in cultural and social factors to shift behaviour.

The United Nations Food and Agriculture Organization and the World Health Organization are meeting this week to discuss the UN Decade of Action on Nutrition. In this issue, we publish two Comment articles that look at some of the problems. On page 30, a group of researchers stresses the importance of nourishing people, not just feeding them. And a piece on page 33 calls for a better approach to quantifying and analysing different aspects of the food-production system.

Between one-fifth and one-third of all food produced goes into the bin. Attention has increased on these post-harvest losses in recent years, and this week the European Commission held what it billed as the first European Union Platform on Food Losses and Food Waste. Launching the event with a speech that will be recognized by any parent who has sat with a child who won't clear their plate, commissioner Vytenis Andriukaitis said: “It is shameful to throw away food in the world where more than eight hundred million people go to bed hungry.”

To help promote responsible development and production of food, the UN global Sustainable Development Goals call for a number of related measures, one of which is to halve waste in the commercial and retail sectors by 2030. It is partly a consumer challenge, because a sizeable chunk of this thrown-away food has been bought and paid for, sometimes at great expense. To many US readers, it will be second nature

to take restaurant leftovers home. That helps to reduce waste, even if some of the contents of the doggy bag end up in, well, the dog.

In countries such as France it's a different story. The government there has been trying to change that, with a new initiative this year that requires restaurants to supply a ‘gourmet bag’ to diners if they ask for one. Not everyone is happy about the idea, and new research offers some pointers why. The results demonstrate, again, the importance of cultural and social factors in shifting behaviour — even in a direction that benefits all concerned.

Writing in the *Journal of Retailing and Consumer Services*, researchers describe interviews with French diners about their attitudes to food waste and taking it home with them (L. Sirieix *et al.* *J. Retail. Consum. Serv.* **34**, 153–158; 2017). They found the now-standard gap between attitudes and behaviour on environmental issues: although three-quarters of respondents thought that doggy bags were a great idea, just one in ten had ever asked for one.

The excuses were many. “It comes from history and French customs,” one said. Another argued that leaving expensive food on the plate showed social status: “Someone who will take home the meal is someone who has less money.” And it was a sign of a downmarket joint: “It will not be well accepted in a fine dining restaurant.”

Diners went further. Asking for leftovers would bring shame. “It's not rude but culturally it's not normal.” And French dishes are just not designed to be eaten that way, they said. It's “too good to be packed” and only acceptable in “a pizzeria”.

One solution, the researchers suggest, is to make the doggy bag desirable: a valuable gift that appeals on more levels than just sustainability. Some high-end restaurants in the United States, they note, package leftovers in a foil swan. That's one possible solution. Another, of course, is simply to serve smaller portions. ■

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

CATALYSTS

Silicon added to life's toolbox

Molecules containing both carbon and silicon have become a standard part of synthetic chemistry, but life uses silicon only in inorganic compounds such as the shells of diatoms.

Now, Frances Arnold and her collaborators at the California Institute of Technology in Pasadena have engineered bacteria to produce carbon–silicon bonds. The team inserted a gene encoding an enzyme from the bacterium *Rhodothermus marinus*, which lives in Iceland's underwater hot springs, into *Escherichia coli*. The resulting bacterium catalysed carbon–silicon bonding in a variety of artificial precursors by inserting carbon into silicon–hydrogen bonds.

The researchers improved the efficiency of their biocatalyst by directed evolution — sequentially inducing mutations in the enzyme's active site and screening for improved activity — and say it could lead to new classes of pharmaceuticals and industrial catalysts using organosilicon molecules.

Science 354, 1048–1051 (2016)
See go.nature.com/2fgot89 for a longer version of this article.

CANCER

'Jet lag' increases mouse cancer risk

Mice with simulated jet lag have an increased risk of developing liver cancer.

Sleeping out of step with the day–night cycle has been linked to various health disorders in humans. David Moore, Loning Fu and their colleagues at Baylor College of Medicine in Houston, Texas, subjected mice to disrupted

day–night cycles using artificial lights to simulate chronic jet lag. Compared with mice under steady 24-hour light–dark cycles, jet-lagged animals had impaired liver metabolism and were more prone to developing fatty liver disease and liver cancer.

Much of the disrupted metabolism was linked to excessive accumulation of bile acid in the liver and altered activities of the nuclear receptors FXR and CAR, which might be targets for preventing liver cancer in future.

Cancer Cell <http://doi.org/btr7> (2016)



IMMUNOLOGY

Macaque social status alters immunity

Changing the social status of macaques also changes their immune systems.

Low social status has been linked to multiple health problems in humans and other primates, some of which may not be linked to the accessibility of food and other resources.

Jenny Tung of Duke University in Durham, North Carolina; Luis Barreiro of the University of Montreal, Canada; and their colleagues analysed the immune cells and gene expression of female macaques before and after they

artificially altered the relative social status of the animals by creating new social groups.

Low status was linked to inflammation, including alterations in the make-up of immune cells in the animals' blood and changes in gene expression that could promote an inflammatory response. This link between social subordination and altered immune response was present despite no variation in access to food or health care among the animals.

Science 354, 1041–1045 (2016)

SEISMOLOGY

Gravity changes before quake hits

Researchers have for the first time detected temporary changes in Earth's gravitational field just before an earthquake.

Jean-Paul Montagnier at the CNRS Institute of Earth Physics in Paris and his team analysed gravimetric and seismic data recorded during the massive 2011 Tohoku earthquake in Japan. They detected a signal that exceeded the background seismic noise and whose amplitude was consistent with that predicted

by theoretical models. The signal occurred several seconds before seismic waves arrived.

Developing sensors to detect such transient gravity signals could lead to new early-warning systems for tsunamis and other earthquake hazards, the authors say.

Nature Commun. 7, 13349 (2016)

CLIMATE CHANGE

Shifting niches spell trouble

Plants and animals will probably be unable to respond to changes in environmental conditions fast enough to keep

pace with climate change.

Tereza Jezkova and John Wiens of the University of Arizona, Tucson, looked at 56 plant and animal species and documented shifts in their 'climatic niche' — the temperature and precipitation conditions they survive in.

They compared these shifts in climatic niches with rates of predicted future change under global warming. On average, expected rates of change in temperature were more than 200,000-fold higher under future climate conditions than rates seen in past shifts, and precipitation changes were more than 10,000-fold higher. Even those groups that had the fastest niche shifts in the past will still experience rates of temperature change that are 300-fold higher in future.

Historical changes were slowest for tropical species, suggesting that they will be especially vulnerable to climate change, the authors warn.

Proc. R. Soc. B <http://doi.org/btn5> (2016)

PALAEOLOGY

Ancient reptile had a diaphragm

The muscle that allows mammals to breathe deeply — the diaphragm — may have been present in some reptiles 300 million years ago, about 50 million years before it was thought to have appeared. Markus Lambertz at the University of Bonn in Germany and his colleagues studied the fossilized skeletons of caseids: extinct mammal-like reptiles that lived 300 million years ago (artist's impression **pictured**). These creatures weighed around 500 kilograms, with short necks that would have made foraging difficult and barrel-shaped bodies that would have restricted their ability to move air into their lungs. Their bones were lightweight and filled with air bubbles, similar to modern whales. The researchers conclude that caseids were probably aquatic, similar to present-day

manatees (sometimes known as sea cows), and must have had muscular diaphragms to inhale all the air they needed during brief trips to the surface.

Ann. N.Y. Acad. Sci. <http://doi.org/btkw> (2016)

QUANTUM PHYSICS

Record set for linked photons

Particles that have linked quantum states, known as 'entangled' particles, can affect each other's states even if they are physically separated. Now scientists have set a record by entangling ten photons — two more than achieved previously.

Entangled particles should one day enable quantum computing and communications, but they are inefficient to produce. A team led by Jian-Wei Pan of the University of Science and Technology of China in Hefei created the ten entangled photons by running five photon pairs through a series of four polarizing beam splitters. They also developed a laser light source that produced their photon batches about 100 times faster than did previous tests.

Phys. Rev. Lett. 117, 210502 (2016)

MATERIALS SCIENCE

Atom-thick coats for copper

Coating metals with a high-quality, single-atom-thick layer of 2D boron nitride can protect them from corrosion.

Layers of 2D materials have been touted for use as protective coatings on surfaces but have shown mixed results,

with some apparently even increasing corrosion rates in the long term.

Zhenguo Huang at the University of Wollongong in Australia and his colleagues used optical microscopes and spectroscopy to study how well films of the insulator boron nitride protect copper from corrosion. They found that copper coated with a high-quality atom-thick film of the material remained almost unaffected by salt water over 1 hour, whereas coating the metal with a multi-layered but cracked film increased corrosion. Such defects probably accelerate oxidation, so coatings could be improved by patching up imperfections and growing films with fewer defects, the authors suggest.

Adv. Mater. <http://doi.org/f3s8tb> (2016)

MATERIALS

Sensor detects UV exposure

A sensor that sticks to the skin can give colour-coded readouts of ultraviolet light levels from the Sun, the leading cause of skin cancer.

John Rogers at the University of Illinois at Urbana-Champaign and his colleagues developed the stretchable device (**pictured**), which is less than one-fifth of a millimetre thick. It has several layers, including a compound that changes the colour of a dye when activated by UV light; filters to monitor both UV-A and UV-B radiation levels; and a small chip that wirelessly sends signals to a nearby device.

The technology could be used to manage UV exposure



on certain regions of the body, and could be adapted to sense other types of environmental exposure, such as pollution and chemical or biological agents, the authors say.

Adv. Funct. Mater. <http://doi.org/f3s7dr> (2016)

EPIGENETICS

Gut microbes shape DNA

Gut microbes and the fatty acids they produce can regulate gene expression by influencing the 3D shape of their hosts' DNA.

Intestinal bacteria are known to affect several aspects of host health, including the risk of cardiovascular and metabolic diseases. To study the mechanisms by which this occurs, Federico Rey and John Denu at the University of Wisconsin-Madison and their colleagues compared mice raised with and without gut bacteria. They found that gut microbes mediate chemical changes to histone proteins, which in turn regulate gene expression by binding to DNA and altering its 3D conformation.

Feeding mice a high-fat, high-sugar diet provides little material for microbes to digest and so blocked some of the changes to DNA shape. Giving short-chain fatty acids (which are produced by gut microbes) to mice raised without gut bacteria restored these effects.

Mol. Cell <http://doi.org/btr6> (2016)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch



SEVEN DAYS

The news in brief

POLICY

Postdoc-pay law

A federal judge in Texas has blocked an effort by the administration of US President Barack Obama to extend overtime pay to millions of US workers. The regulation sought to double to US\$47,500 the salary cut-off below which workers would be eligible for mandatory overtime pay. Set to come into force on 1 December, it would have raised pay for many US postdocs, whose average salary is around \$45,000, with many earning substantially less. On 22 November, the judge agreed with 21 states and a coalition of business groups that the rule is unlawful, and issued a nationwide injunction.

PHARMACEUTICALS

Alzheimer's drug

A drug that was a leading hope in the fight to treat Alzheimer's disease has failed in phase III clinical trials, its developer announced on 23 November. Solanezumab, made by Eli Lilly of Indianapolis, Indiana, did not slow cognitive decline in people with Alzheimer's compared with those who took a placebo. The drug is an antibody-based treatment that targets amyloid protein in the brain. The trial was seen as a major test of the 'amyloid hypothesis', which posits that a build-up of the protein triggers Alzheimer's. The news sent stock in Eli Lilly, which invested heavily in the therapy, tumbling. See page 15 for more.

Cancer-trial deaths

A promising but risky cancer therapy has again led to patient deaths, prompting the treatment developer Juno Therapeutics to halt a key clinical trial. The company,

located in Seattle, Washington, announced on 23 November that two patients had died of severe brain swelling during a clinical trial of an immune therapy. The treatment involves removing a person's T cells, engineering them to recognize leukaemia cells, and reinjecting them. The death of three other patients this year led to a hold on the trial, which was lifted in July after the study protocol was changed.

HIV-vaccine trials

A large-scale clinical trial of two promising HIV vaccines has begun in South Africa, health officials announced on 27 November. The study

is a long-awaited follow-up to a trial of a vaccine in Thailand that started in 2003 and reduced the risk of HIV infection by 31%. It will test two experimental vaccines similar to the one used in the Thai trial, and will aim to enrol 5,400 people at 15 sites in South Africa. Results on the effectiveness of the vaccines are expected in 2020.

PUBLISHING

Boycott threat

More than 2,100 people have signed a petition coordinated by Finnish scientists calling on major scientific publishers to make

for Reconstruction and Development, the project's main funder. The US\$1.6-billion shield, called the New Safe Confinement, should secure the reactor unit for 100 years and contains remotely controllable cranes and machinery to dismantle the ageing concrete shelter that it replaces. Installation was completed on 29 November.



SERGEY DOLZHENKO/EPA

Huge safety shield erected at Chernobyl

A massive radiation shield has been installed over the number four nuclear reactor in Chernobyl, Ukraine, the site of the catastrophic 1986 nuclear accident. The steel-and-concrete structure is 275 metres wide, 108 metres high and weighs 36,000 tonnes, and is the largest movable land-based structure ever built, according to the European Bank

subscription prices "more reasonable" and to increase open access to research content. The researchers are supporting demands made by FinElib, a consortium of Finnish libraries, which is negotiating contracts for 2017 subscriptions with several large publishers. Two-thirds of signatories have said that they would be willing to boycott the journals involved and abstain from peer-review duties if the demands are not met. Last year, Finnish research organizations paid €27 million (US\$28.6 million) for journal subscriptions, says FinElib, but flat or falling budgets cannot support rising expenditure.

LEON NEAL/GETTY The petition follows a stand-off last year between universities in the Netherlands and the Dutch publisher Elsevier, which ended in a deal to make some papers by Dutch scientists open access.

EVENTS

Women of science

More than 10,500 female scientists have signed an open letter that pledges to combat the “divisive and destructive rhetoric” and anti-science sentiment that has followed the US election. A group of scientists initially wanted to build a network of 500 women to start a positive conversation at a time when they felt that science and their social values were under attack — but within days, thousands had signed up. The group now hopes to develop into a global network for research support and to inspire young women to pursue science.

Top-price dodo

A rare, near-complete skeleton of a dodo (pictured) sold at auction in Britain for £280,000 (US\$350,000) on 22 November. The bones — which come from a number of specimens but together form a 95%-complete skeleton — were compiled by an enthusiast over four decades. Few traces remain of dodos (*Raphus cucullatus*), flightless



birds that were found on Mauritius and were driven rapidly to extinction in the seventeenth century. There are reportedly only 12 similarly complete skeletons in the world, as well as a few other head and foot bones.

CLIMATE

EU climate target

The European Union will probably fail to reach its target of spending one-fifth of its €1-trillion (US\$1-trillion) budget on climate action in 2014–20, the European Court of Auditors in Luxembourg said on 22 November. According to the European Commission, the proportion spent on climate action averaged 17.6% in 2014–16. The EU would need to increase its spending to 22% for 2017–20, which the auditors think is unlikely — they found no shift towards higher spending on climate action in agricultural, rural-development or fisheries policies. The Horizon 2020

research-funding programme is also falling behind on its target, having allocated only 24% instead of 35% of its budget to climate research for 2014–17.

Worst reef die-off

Surveys of Australia's Great Barrier Reef in the wake of this year's huge 'bleaching' event have confirmed that 67% of shallow-water corals have died in a 700-kilometre stretch of the reef's northern region. When water is too warm, corals expel symbiotic algae and turn white. High water temperatures also caused 26% of corals to die in a region farther north, but the central and southern regions were less affected. This bleaching event — the third in the reef's recorded history — is the worst yet, say researchers from the Australian Research Council Centre of Excellence for Coral Reef Studies at James Cook University in Townsville.

SPACE

Negative altitude

The European Space Agency on 23 November revealed more details from its investigation into why its lander Schiaparelli crashed on Mars in October. The probe was part of the ExoMars mission and was designed to test landing technology. The latest information

COMING UP

3–6 DECEMBER

The American Society of Hematology holds its annual meeting in San Diego, California.
go.nature.com/2goxht8

4–17 DECEMBER

The 13th conference of the parties to the Convention on Biological Diversity meets in Cancun, Mexico.
www.cbd.int/cop

suggests that there was a glitch in its rotation sensors — and confirms a previous suggestion that Schiaparelli ran into trouble when the sensor data were fed into its navigation system. The issue caused the lander to think it was at a lower altitude than it really was, and to jettison its parachute too early. The new twist: Schiaparelli not only thought it was lower than it was — it thought it was below ground level.

Star names

The International Astronomical Union (IAU) has established a new catalogue of star names in an effort to formalize the colloquial names often given to stars. Astronomers tend to use alphanumeric designations for stars, but many traditional names do not have official spellings, and sometimes the same name may be used for different stars. On 24 November, the IAU announced that it had approved a set of 227 designations, decided by its Working Group on Star Names, which it established in May. Among them are Fomalhaut, Sirius and Rigel Kentaurus, the ancient name for Alpha Centauri, our closest star system.

NATURE.COM

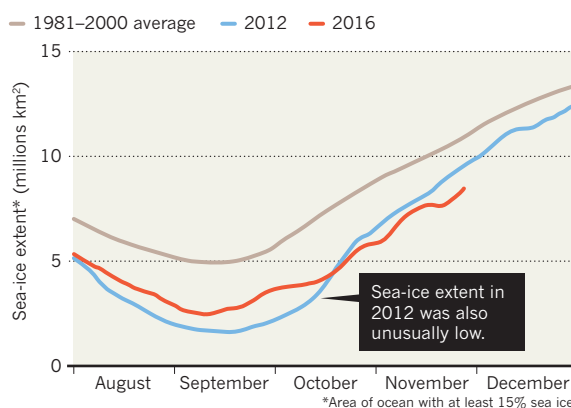
For daily news updates see:
www.nature.com/news

TREND WATCH

Scientists have again raised the alarm about the increasingly fragile state of the Arctic. The *Arctic Resilience Report*, published on 25 November, highlights how ongoing changes, such as shrinking sea ice and melting permafrost, could interact in ways that further harm ecosystems and the people who depend on them. The report comes as Arctic sea-ice cover is lower than usual during the winter freeze-up (see chart). This October's sea-ice extent was the lowest for any October average in the satellite record.

LOW ARCTIC SEA ICE WORRIES SCIENTISTS

Warm sea and air temperatures in the Arctic have contributed to the lowest extent of sea ice ever recorded for the time of year.



NEWS IN FOCUS

GEOSCIENCE Speedy Antarctic drill kicks off hunt for Earth's oldest ice **p.18**

DRUG DISCOVERY Weaponized antibodies upgraded to fight cancer **p.19**

BIOTECHNOLOGY Gene-editing controversy enters peer-reviewed literature **p.20**

ENVIRONMENT Cleaning up the e-waste industry in the West Bank **p.22**



JODI COBB/NATIONAL GEOGRAPHIC CREATIVE



A woman comforts her twin, who has Alzheimer's disease — the underlying cause of the condition is still a matter of debate.

PHARMACEUTICALS

Leading Alzheimer's theory survives drug failure

Solanezumab flopped in a large clinical trial, but the drug or others like it could yet succeed.

BY ALISON ABBOTT & ELIE DOLGIN

A drug that was seen as a major test of the leading theory behind Alzheimer's disease has failed in a large trial of people with mild dementia. Critics of the 'amyloid hypothesis', which posits that the disease is triggered by a build-up of amyloid protein in the brain, say the results are evidence of its weakness. But the jury is still out on whether the theory will eventually yield a treatment.

Proponents of the theory note that the failure could have been due to the particular

way in which solanezumab, the drug involved in the trial, works, rather than a flaw in the hypothesis. And other trials are still ongoing to test whether solanezumab — or other drugs that target amyloid — could work in people at risk of the disease who have not shown symptoms, or even in people with Alzheimer's.

"I'm extremely disappointed for patients, but this, for me, doesn't change the way I think about the amyloid hypothesis," says Reisa Sperling, a neurologist at the Brigham and Women's Hospital in Boston, Massachusetts.

She is leading one of several trials to test

whether drugs that aim to reduce the build-up of amyloid 'plaques' can prevent Alzheimer's in people at risk of the disease.

Solanezumab is an antibody that mops up amyloid proteins, which can go on to form plaques in the brain, from the blood and cerebrospinal fluid. Eli Lilly, the company that developed the drug, announced on 23 November that it would abandon it as a treatment for people with mild dementia. The outcome adds to a long list of promising Alzheimer's drugs that have flopped in the clinic, many of which, like solanezumab, targeted amyloid. ▶

► The Lilly trial tracked more than 2,100 people diagnosed with mild dementia due to Alzheimer's disease for 18 months. Half received monthly infusions of solanezumab, the other half a placebo. Analysis of people with comparable symptoms in earlier studies of solanezumab had seemed encouraging, but the latest trial indicated only a small cognitive benefit, not enough to warrant marketing the drug (see 'Market reaction').

PREVENTION HOPE

Lilly has also been running prevention trials to see whether solanezumab might help people at especially high risk of the disease. The company says it will now discuss with its trial partners whether to continue with those.

Sperling's trial is one of these, and tests solanezumab in people who have elevated amyloid levels in the brain but have not shown any symptoms of dementia. Researchers at Washington University in St Louis, Missouri, are also testing solanezumab, and a similar antibody made by drug company Roche, in people who are currently healthy but are genetically at high risk of developing Alzheimer's. Meanwhile, the Banner Alzheimer's Institute in Phoenix, Arizona, is testing the effects of three therapies that target amyloid production, one of which is an antibody, in people at high genetic risk of Alzheimer's. The Lilly outcome "doesn't disprove the

MARKET REACTION

When company Eli Lilly announced that an Alzheimer's drug based on the amyloid hypothesis had flopped in a clinical trial, its stock plummeted.



amyloid hypothesis, and it really increases the importance of these longer prevention trials", says Eric Reiman, the Banner Institute's executive director and leader of the trials.

Lilly's result may say more about the characteristics of solanezumab than about the accuracy of the underlying amyloid hypothesis, says Christian Haass, head of the Munich branch of the German Centre for Neurodegenerative Diseases. The antibody targets soluble forms of amyloid, he points out, so it "could be trapped in the blood without ever reaching the

actual target in the brain in sufficient quantities".

Biogen, a company based in Cambridge, Massachusetts, is testing a different antibody called aducanumab, which targets amyloid plaques in the brain. In early clinical testing, the antibody showed signs of clearing amyloid and alleviating memory loss in people with mild Alzheimer's disease; results from phase III trials are expected in 2020.

"Until the aducanumab data read out, we have not truly put amyloid to the test," says Josh Schimmer, a biotechnology analyst at financial-services firm Piper Jaffray in New York City.

Still, the negative trial findings have emboldened critics of the amyloid theory, who are weary of its failure to yield a treatment. "The amyloid hypothesis is dead," says George Perry, a neuroscientist at the University of Texas at San Antonio. "There's no sign of anybody getting better, even for a short period, and that suggests to me that you have the wrong mechanism," adds Peter Davies, an Alzheimer's researcher at the Feinstein Institute for Medical Research in Manhasset, New York.

No matter what Lilly decides about its other solanezumab trials, the company isn't giving up on Alzheimer's. It is testing an inhibitor of an enzyme involved in the synthesis of amyloid in partnership with AstraZeneca, and is progressing with a handful of candidate therapies aimed at other targets. ■

SOURCE: NYSE

POLITICS

Bonus funds excite UK scientists

Government announces extra £4.7 billion for research and development up to 2021.

BY ELIZABETH GIBNEY

British scientists are not used to hearing about large increases in national research spending. So when Prime Minister Theresa May promised on 21 November that her government would invest an extra £2 billion (US\$2.5 billion) per year in research and development (R&D) by 2020, scientists gave her speech a cautious welcome.

But the funding hike seems to be no

financial sleight of hand, according to Treasury documents released on 23 November after Chancellor of the Exchequer Philip Hammond gave an address on the nation's finances. The government expects to spend an extra £4.7 billion on R&D between now and 2020–21, it says, and the final year's £2-billion boost will represent a rise of around 20% in total government R&D spending. Still, it remains unclear how the cash will be allocated, and how much will fund basic, blue-skies research.

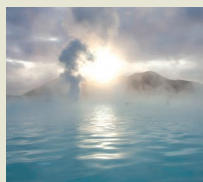
"It is a real boost to see UK strength in science being championed by the prime minister and backed with what is the most significant investment in R&D I can remember," said Sarah Main, director of the London-based Campaign for Science and Engineering.

INDUSTRIAL CHALLENGES

Some of the money will go directly to applied R&D through a new Industrial Strategy Challenge Fund, modelled on the US Defense



TOP NEWS



Living cells bind silicon and carbon for the first time
go.nature.com/2fgo189

MORE NEWS

- External brain stimulation goes deep go.nature.com/2gepapp
- Peer-review 'heroes' do lion's share of the work go.nature.com/2fhosc6
- Tracking the transition: science agencies under Donald Trump go.nature.com/2gdtbho

NATURE PODCAST



CRISPR's rival stumbles; Pluto's icy heart; and is mitochondrial replacement ready for the clinic? nature.com/nature/podcast

ARNALDUR HALDORSSON/
BLOOMBERG VIA GETTY

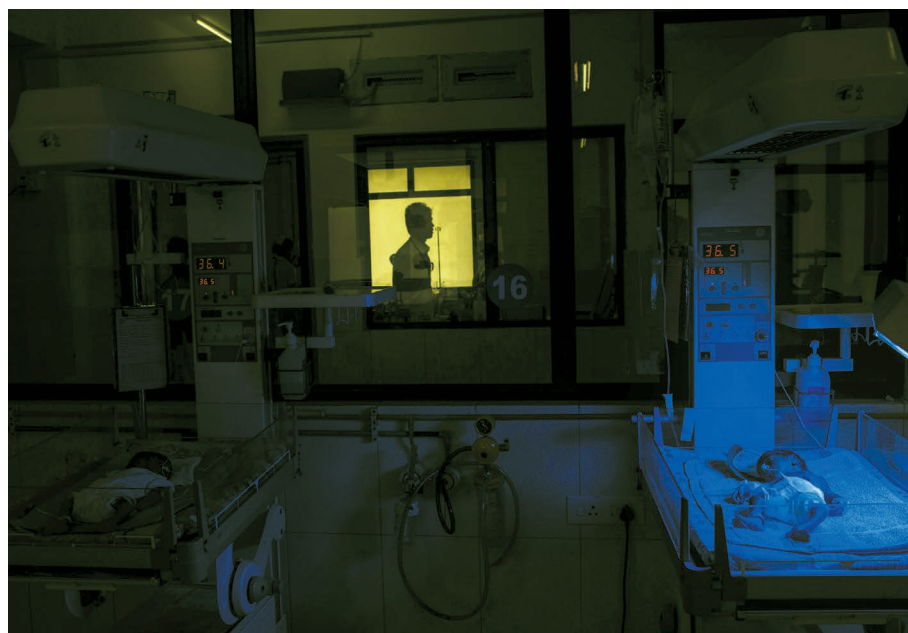
Advanced Research Projects Agency (DARPA), the Pentagon's high-risk research arm. That fund will support cross-disciplinary "collaborations between business and the UK's science base", according to the Treasury documents, and will "set identifiable challenges for UK researchers to tackle". It will be managed by Innovate UK, a government body that funds R&D primarily through businesses, and by the seven UK research councils, which mainly fund university research. The money will be allocated using an "evidence-based process", the Treasury says. Other cash will go to "innovation, applied science and research". Although the Treasury was vague on what this entailed, it said that the funding would be used "to increase research capacity and business innovation" and "to further support the UK's world-leading research base". UK Research and Innovation (UKRI) — an agency that has not yet been created, but is expected to unite the research councils and Innovate UK — will award the funding on the basis of "national excellence".

The documents make no clear reference to spending on basic research, but Main said she would be surprised if it was excluded. "I think it will be really important that this funding goes to both blue-skies and challenge-driven research. It is clear from the document that there is money there just to increase the UK's research capacity, and that this money is going to be channelled through UKRI. It will be important for UKRI to consider the balance of how that money is distributed," she said.

BREXIT AHEAD

Scientists had hoped that the speech would signal their new government's approach to science. And although it was dominated by forecasts of the country's slowed economic growth as a result of Brexit — its exit from the European Union — research and innovation also took top billing. "We do not invest enough in research, development and innovation," Hammond said.

But "sorely missing" from the statement was any reference to the impact on science from Brexit, says Stephen Curry, a structural biologist at Imperial College London and a member of the advisory board for the campaign group Science is Vital. "I'd like to know how the loss of EU funding will impact decisions on allocation of the new investments," Hammond's opposition counterpart, Labour Party shadow chancellor John McDonnell, replied by saying that the rise in R&D funding was not enough: it would lift the proportion of UK gross domestic product spent on R&D from 1.7% to only 1.8%. The Organisation for Economic Co-operation and Development recommends that developed countries should be spending 3%. ■



SANJIT DAS/PANOS

Physicians may soon have a lot more help in treating newborns.

MEDICINE

Preventing brain damage in babies

Experimental therapies could save thousands of newborns.

BY ERIKA CHECK HAYDEN

Neuroscientists and physicians have embarked on what they hope will be a revolution in treatments to prevent brain damage in newborn babies.

As many as 800,000 babies die each year when blood and oxygen stop flowing to the brain around the time of birth. And thousands develop brain damage that causes long-lasting mental or physical disabilities, such as cerebral palsy. Physicians have few tools to prevent this, but they are optimistic that clinical trials now under way will change things.

The trials were sparked by neuroscientists' realization in the 1990s that some brain injuries can be repaired. That discovery spurred a flurry of basic research that is just now coming to fruition in the clinic.

In January, a US study will start to test whether the hormone erythropoietin, or EPO, can prevent brain damage hours after birth when combined with hypothermia, in which babies are cooled to 33.5°C. A trial in Australia is already testing this treatment. Physicians in countries including the United States, China and Switzerland are testing EPO in premature babies, as well as other treatments,

such as melatonin, xenon, argon, magnesium, allopurinol and cord blood in full-term babies.

"The world has really changed for us," says neurologist Janet Soul at Boston Children's Hospital in Massachusetts.

Therapeutic hypothermia was the first success: clinical trials over the past decade have shown that it decreases the risk of death and of major brain-development disorders by as much as 60%. It is now standard treatment for babies in developed countries whose brains are deprived of blood and oxygen during birth.

"I can't tell you how great it was to be able to do something for these babies."

"I can't tell you how great it was to be able to do something for these babies rather than stand there and watch them have seizures," Soul says.

But because hypothermia does not work for all babies¹, scientists decided to see whether combining it with other treatments would help. EPO was known to boost the production of red blood cells even before its discovery² in mouse brain cells in 1993, and is regularly used by physicians to treat anaemia. Neuroscientist Sandra Juul at the University of ▶

► Washington in Seattle wondered what a blood-boosting hormone was doing in the brain. In subsequent animal studies, she found that the hormone stopped brain cells from dying and helped the brain to repair itself³. That led a few years later to the first clinical trials showing that EPO prevents brain damage in babies.

MOVING ON

In June, a study conducted by Juul and her colleagues reported the results of giving EPO or a placebo, along with inducing hypothermia, just after birth to dozens of babies at risk of brain injury. Those who received EPO were less likely than those given the placebo to show signs of brain damage on magnetic resonance imaging tests done five days later⁴.

Those results led to the forthcoming clinical study. Co-led by Juul and Yvonne Wu, a paediatric neurologist at the University of California, San Francisco, the trial will enrol 500 babies at risk of brain injury from 17 hospitals across the United States during their first 24 hours of life.

All the babies will be treated with hypothermia. Half will then receive five doses of EPO over seven days; the other half will get saline injections. The US\$10-million trial will measure whether the hormone boosts the children's mental and physical health at 2 years of age.

Researchers are also testing EPO in babies born as early as 23 weeks in the United States and Europe. Such premature babies are more likely to develop brain injury than are full-term babies, and smaller studies have produced conflicting results about the benefits of EPO in these very early cases.

But neonatologist Giancarlo Natalucci of the University of Zurich, who was part of a Swiss trial that found EPO didn't improve the health of two-year-olds who had been treated as premature babies⁵, says that factors such as dose may account for such results. He still thinks that the treatment merits study.

The trials are difficult to conduct because it's hard to tell whether a symptom is a side effect of treatment or the result of a baby's underlying injuries.

But despite the hurdles, Juul and other researchers press on, driven by their desire to aid the world's smallest patients. "They're in such desperate need of help," Juul says. ■

1. Azzopardi, D. et al. *N. Engl. J. Med.* **371**, 140–149 (2014).
2. Masuda, S. et al. *J. Biol. Chem.* **268**, 11208–11216 (1993).
3. Kellert, B. A., McPherson, R. J. & Juul, S. E. *Pediatric Res.* **61**, 451–455 (2007).
4. Wu, Y. W. et al. *Pediatrics* **137**, e20160191 (2016).
5. Natalucci, G. et al. *J. Am. Med. Assoc.* **315**, 2079–2085 (2016).

ANTARCTICA

Speedy drills start hunt for oldest ice

British team first to seek site of 1.5-million-year-old sample.

BY QUIRIN SCHIERMEIER

As the short Antarctic spring ends and long summer days approach, geoscientists are flocking to the frozen continent to start a new kind of exploration.

In December, the first drill designed to search for a scientifically useful sample of ice that is at least 1.5 million years old will begin its work. It is part of a broader effort to locate the best place to extract a core containing Earth's oldest ice, which would help to reveal how climate has shaped the planet's past and how to predict future fluctuations.

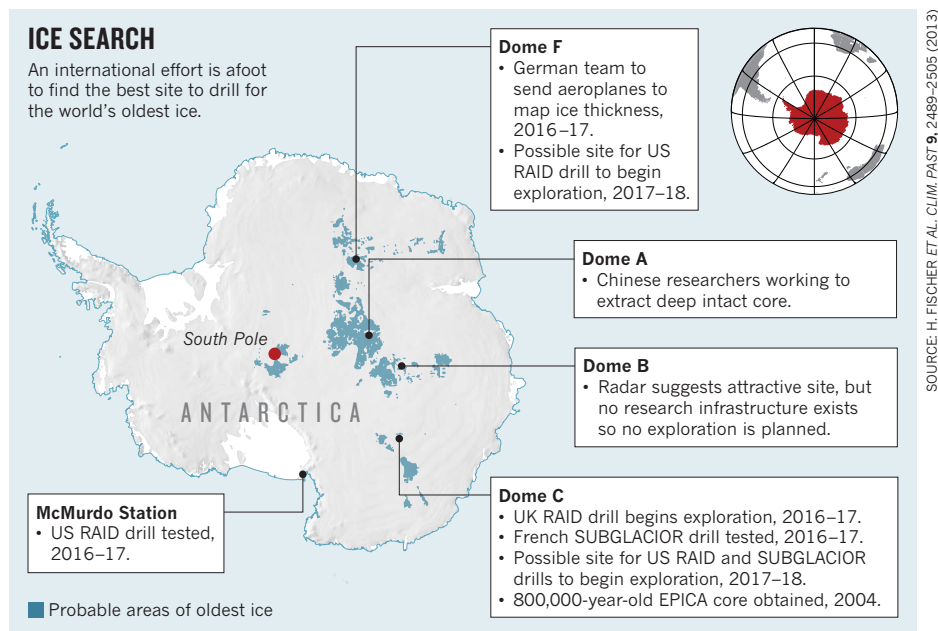
"This exciting field season should bring us a large step nearer to deciding where to drill the oldest-ice core," says Olaf Eisen, a glaciologist at the Alfred Wegener Institute of Polar and Marine Research in Bremerhaven, Germany, who coordinates an exploration team funded by the European Union.

More than a decade ago, the European Project for Ice Coring in Antarctica (EPICA) drilled the oldest existing core, which contains 800,000-year-old ice, from an ice dome in East Antarctica known as Dome C. The core reaches only as far back as the latter part of the Pleistocene epoch, when Earth began cycling between warm and cold periods every 100,000 years.

Before 1 million years ago, the cycle occurred every 40,000 years (L. E. Lisiecki and M. E. Raymo *Paleoceanography* **20**, PA1003; 2005), so scientists want an ice core that is twice as old as EPICA to better understand this transition.

Digging such a core would cost about US\$50 million and take several years, so researchers want to be sure that the location is optimal — with ice that is sufficiently deep but not melted at the bottom by geothermal activity. "It's absolutely crucial to thoroughly investigate all options," says Eisen. Enter a new breed of drill, designed to do fast, cheap reconnaissance instead of extracting a single, intact ice core, as previous deep drills have done.

One promising location, 'little Dome C', lies just 40 kilometres away from the EPICA site — and is where the £500,000 (US\$620,000) Rapid Access Isotope Drill (RAID) will start boring this month, led by climate scientist Robert Mulvaney of the British Antarctic Survey in Cambridge, UK. A narrow drill, RAID will excavate to 600 metres in about 7 days — compared with 5 years for a 3.4-kilometre core such as EPICA's. And rather than extract a core, RAID will measure the ice's temperature and collect chips of ice. Scientists will then comb these for clues from isotopes as to the age and temperature of the ice at the bottom of the sheet.



A more expensive reconnaissance instrument, which will do its first drilling tests this season at Dome C, should be ready to take a deeper look next year. Led by glaciologist Jérôme Chappellaz of Joseph Fourier University in Grenoble, France, the €3.2-million (US\$3.4-million) SUBGLACIOR probe, which is about the same width as RAID, can penetrate the more than 3-kilometre-thick ice sheet in a single season.

Both the UK and French drilling projects are funded as part of the EU collaboration. But also this season, a US team led by climatologist Jeffrey Severinghaus of the Scripps Institution of Oceanography in La Jolla, California, and John Goodge of the University of Minnesota, Duluth, will test the \$10.5-million Rapid Access Ice Drill (also abbreviated RAID) at Minna Bluff, near the US McMurdo Station on Ross Island. Producing a hole of about 8 centimetres — similar to the boreholes of the other drills — it is the only rapid drill that can extract rocks from the bottom of a core. Next Antarctic summer, the

team will begin its hunt for the site of a 1.5-million-year-old core.

Dome C is one option for its first excavation. Another is the relatively unexplored Dome F in Antarctica's Queen Maud Land, which ground-based radar suggests is a promising candidate (see 'Ice search'). In January, a German team will run reconnaissance flights there. Funded by the same European grant as the UK RAID and SUBGLACIOR drills,

this radar survey will give a more comprehensive view of the ice thickness. Severinghaus says that his team will watch for the data when deciding where to point the US RAID.

Both the US and European teams are working under an umbrella group. The International Partnerships in Ice Core Sciences (IPICS) aims to identify a suitable site to drill a core representing Antarctica's oldest ice in the next two years. That drilling could start by the end of 2020, says Eisen. But how the

international teams would work together on a joint project, or share funding, is unclear.

There's a possibility that a record-breaking ancient core could show up sooner. For several years, scientists at the Polar Research Institute of China in Shanghai, who are also members of IPICS, have been probing the ice sheet that covers Dome A, a plateau close to the centre of the Antarctic continent. Using a conventional corer rather than a rapid exploratory drill, they are working on obtaining a deep, intact ice core from the region, says Eisen — and it is possible that it could stretch back to 1.5 million years. Such a surprise success would increase pressure on teams from other nations to produce their own record, he says.

Multiple cores would benefit science. "We would carry on with our project," says Eisen. The IPICS effort would ideally excavate multiple 1.5-million-year-old cores in any case. "You cannot trust a single core," says Severinghaus. "We absolutely need different records from different thermal regimes." ■

DRUG DISCOVERY

Weaponized antibodies use new tricks to fight cancer

Next generation of Trojan-horse drugs designed to minimize damage to healthy cells.

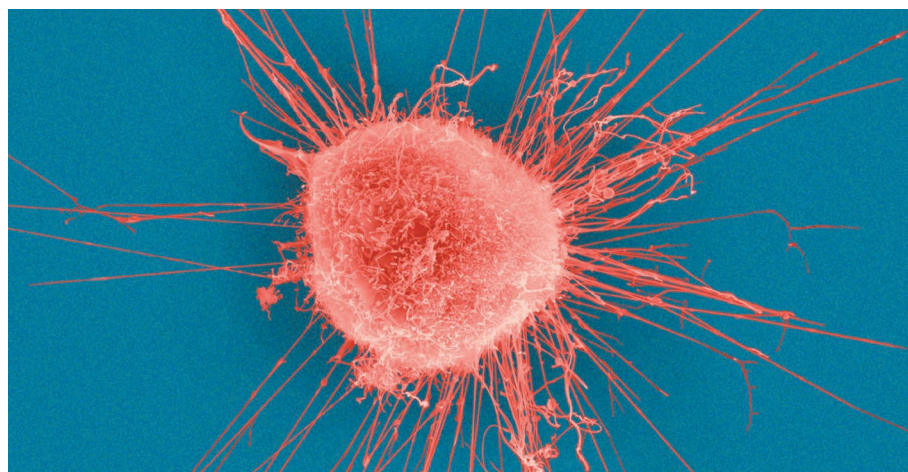
BY HEIDI LEDFORD

After decades of frustration, efforts to develop antibodies that can ferry drugs into cancer cells — and minimize damage to healthy tissue — are gathering steam. The next generation of these 'weaponized antibody' therapies, called antibody-drug conjugates (ADCs), is working its way through clinical trials.

Researchers will gather to discuss this renaissance on 30 November at the Symposium on Molecular Targets and Cancer Therapeutics in Munich, Germany. The improvements come after the first wave of experimental ADCs failed to deliver on its promise.

"Initially there was a lot of excitement, and then slowly many of them did not work," says Raffit Hassan, a cancer researcher at the US National Cancer Institute in Bethesda, Maryland. Now, he says, there are two new ADCs in phase III clinical trials, and many more in earlier-stage testing.

The concept that underlies these drugs is simple: repurposing an antibody as a vehicle to deliver a toxic drug into a cancer cell. When the antibody in an ADC seeks out and docks



DENNIS KUNKEL MICROSCOPY/SPL

Breast-cancer cells might be susceptible to drugs that masquerade as antibodies to sidestep defences.

onto a tumour cell, the cell takes it up and cleaves the molecular links that bind the drug to the antibody. This frees the drug to kill the cell from within.

But this approach has proved tricky to realize. Sometimes the molecular linkers are too tight, and do not release the drug inside the cell. Sometimes they are too unstable, and release the drug near healthy cells — limiting

the dose that can be administered. Even the drugs themselves can be problematic: because most are toxic mainly to rapidly dividing cells, they can leave behind the slowly dividing cells that seed some tumours. And some have had trouble penetrating more than a few cell layers into their target tumours.

Researchers have been chasing ADCs for decades, Hassan says. The US Food and

► Drug Administration has approved three, but one was subsequently withdrawn from the market amid concerns that it was not effective and posed safety risks. The other two have met a happier fate: sales of Adcetris (brentuximab vedotin), approved in 2011 to treat lymphoma, and Kadcyla (trastuzumab emtansine), approved in 2013 to treat breast cancer, have been encouraging, says Ryan Million, head of the San Francisco office of the life-sciences and health-care consultancy firm Trinity Partners.

The approvals gave investors confidence in the field and sent researchers into a frenzy to improve their designs. More than 40 ADCs are now in clinical testing. Genentech, the biotechnology firm in South San Francisco, California, that developed Kadcyla, is experimenting with alternative drugs and molecular linkers. “Chemistry efforts have gotten more sophisticated in making decisions about which linker will go with each drug,” says Bernard Fine, a group medical director at the firm. The company is now working on nine ADCs.

Researchers are also mining a wealth of data from cancer-sequencing projects in search of new targets for antibodies to latch onto, says Stéphane Depil, medical director of the cancer immunotherapy programme at the Centre Léon Bérard in Lyon, France. Identifying those that are unique, or nearly so, to cancer cells has been a major challenge, he says. But growing interest in harnessing the immune system has led researchers to catalogue unique proteins expressed on the surface of malignant cells.

Some companies are trying to hit familiar targets with entirely new designs. Mersana Therapeutics, a biotechnology firm in Cambridge, Massachusetts, has attached both an antibody and a drug to a biodegradable polymer, rather than linking them to each other. This allows the company to attach 15 molecules of the drug to each polymer, rather than the usual three or four, says chief scientific officer Timothy Lowinger. Mersana is testing its approach in early clinical trials of a drug conjugate that targets HER2, a protein expressed at high levels in some breast-cancer tumours. Kadcyla targets HER2, too, but Lowinger says that Mersana’s version can bring in more drug per target, so it could be useful against cancers that express only low levels of HER2.

And at Tarveda Therapeutics, a biotechnology company in Watertown, Massachusetts, researchers have dispensed with the antibody altogether. Instead they are using a short strand of amino acids, the building blocks of proteins, to target cancer cells. The result is a drug that is about 15 times smaller and likely to penetrate deeper into the tumour, says Richard Wooster, Tarveda’s president of research and development.

Even with all this activity, the technology has not reached its peak, says Million. “There’s still lots to innovate,” he says. “But when it works, I think it will work powerfully.” ■



Zebrafish embryos star in one peer-reviewed paper investigating the NgAgo gene-editing technique.

BIOTECHNOLOGY

Gene-editing row escalates

Attempts to use a controversial potential rival to the CRISPR–Cas9 technique have now been published.

BY DAVID CYRANOSKI

A heated dispute over gene-editing that began in blogs and social media is now playing out in the scientific literature.

Six months ago, Chinese researchers reported that an enzyme called NgAgo could be used to edit mammalian genes¹ — and that it might be more accurate and more versatile than the popular CRISPR–Cas9 gene-editor. But other scientists complained that they could not replicate the experiment.

Now, a paper published in *Protein & Cell*² lists multiple failed replications, and another, published in *Cell Research*³, suggests that NgAgo may only block, but not edit, genes when injected into zebrafish (*Danio rerio*). *Nature Biotechnology*, which published the first NgAgo paper, has also published a report of three failed attempts to replicate the original experiment⁴, and an ‘expression of concern’ to accompany the original paper.

Nature Biotechnology is editorially independent of *Nature*’s news team and is owned by *Nature*’s publisher, Springer Nature.

Han Chunyu, a biologist at Hebei University of Science and Technology in Shijiazhuang who first reported the NgAgo experiment, says that he stands by his team’s original claims and that “the *Nature Biotechnology* paper provides us some clues as to why others are having problems.” He says that he hopes to submit a scientific paper explaining why others are having difficulty by the end of the year.

Nature Biotechnology says that it will give Han’s team the opportunity to respond to the criticisms in the report by January 2017. “An update will be provided to the community at that time,” said a spokesperson.

Gene-editing techniques that precisely disable or modify specific sections of a genome have taken the biomedical world by storm. NgAgo is one of several proposed alternatives to the most popular method, CRISPR–Cas9.

The 20 authors of the *Protein & Cell* paper² describe how they attempted without success to use NgAgo to edit a variety of genomes. Eight of the labs then tried the feat again, using genetic materials provided by Han, targeting the same genes and also applying the technique to human cells. They all failed.

The paper urges Han's team to "clarify the uncertainty surrounding NgAgo". One author, Wensheng Wei, a molecular biologist at Peking University, has already made up his mind. "It simply doesn't work, period," he says.

Zhang Xiaoxue, managing editor at *Protein & Cell* in Beijing, says that the journal made an effort to publish the NgAgo paper quickly because of the debate. "In China, it's not just a scientific issue. It's also an ethical and political issue," she says.

The failed replications described in *Nature Biotechnology*⁴ were carried out by three more groups; all also used genetic materials provided by Han, targeted the same genes and applied the technique to human cells.

In the *Cell Research* paper³, researchers report an attempt to use NgAgo to edit a gene in zebrafish embryos that is thought to be related to eye development. Some of the embryos developed either one very small eye and one largely normal eye, or eyes that were fused and formed, according to the paper, on the top of the head "like a cyclops", as if NgAgo had knocked out the gene. But sequencing the genomes of the fish revealed that the gene was still intact.

Lead author Liu Dong, a molecular biologist at Nantong University in China, suggests that the NgAgo molecules clamp onto the genome, but instead of cutting the target gene, they reduce its expression. If he is right, then

NgAgo does not make permanent changes that are passed on to the next generation and would therefore not be considered a gene editor.

But Liu offers little insight into the controversy over the original NgAgo experiments, which, he notes, were done in human cells *in vitro*. He adds that the NgAgo protein, which can be easily prepared in the laboratory, could provide a cheap, accessible alternative to current methods of temporarily blocking gene function in zebrafish.

"This is another report confirming that NgAgo does not work as a gene editor."

Confirming that NgAgo does not work as a gene editor," says Lluís Montoliu, a geneticist at the Spanish National Centre for Biotechnology in Madrid, who has previously criticized the Han paper.

Han says that he has found a problem that could explain why others are having difficulty replicating his results. "I cannot say right now because the media in China jumps on everything I say," he says. "I need a little bit of time."

One of the few scientists who previously told *Nature* that he had corroborated Han's

findings now says that he is using NgAgo, and that he hopes to publish soon. But another who previously noted positive results with NgAgo says now that the "data are confusing" and "we cannot make a conclusion". Neither researcher wanted to be named, for fear of being dragged into the controversy.

The debacle has raised questions about a 224-million-yuan (US\$32-million) gene-editing centre that Han's university — the Hebei University of Science and Technology — announced in August that it will build. "Without Han's *Nature Biotechnology* paper and the hype after that, it's impossible for the school to get such huge funding," says Fang Shimin, a former biochemist who has become famous for exposing fraudulent scientists. He was also one of the first to publicize criticism of Han's paper. If Han's work doesn't stand up, "the centre will lose its legitimacy," he says. The university declined requests to discuss the centre. ■

1. Gao, F., Shen, X. Z., Jiang, F., Wu, Y. & Han, C. *Nature Biotechnol.* **34**, 768–773 (2016).
2. Burgess, S. et al. *Protein Cell* <http://dx.doi.org/10.1007/s13238-016-0343-9> (2016).
3. Qi, J. et al. *Cell Res.* <http://dx.doi.org/10.1038/cr.2016.134> (2016).
4. Lee, S-H. et al. *Nature Biotechnol.* <http://dx.doi.org/10.1038/nbt.3753> (2016).

See <http://go.nature.com/2fujd8m> for a longer version of this story.

CORRECTION

In the story 'Speedy drills start hunt for oldest ice' (*Nature* **540**, 18–19; 2016), the size of the borehole to be drilled by the US RAID was wrong. It will be about 8 cm, and so roughly the same size as the other planned holes. The story should also have noted that as well as Jeffrey Severinghaus, the US project is co-led by John Goodge of the University of Minnesota, Duluth.

The ghosts of old computers

In a rare example of cooperation, Israelis and Palestinians are working together to clean up electronic waste in polluted communities.

By Josie Glausiusz

Bidoo Cave in the hills west of Hebron opens with towering arches that lead through subterranean chambers carved perhaps thousands of years ago. Cut into the walls of one cavern are hundreds of small, rectangular niches where ancient residents once raised pigeons for meat, eggs and ritual sacrifice.

The cave could be an archaeological treasure, but soot coats the walls and the floor is littered with rubbish, including burnt tyres and wiring. Deep inside the cavern, beside a giant rock, is a lone upside-down computer monitor.

Bidoo is used by local children to burn ‘e-waste’ — mostly leftover foams and plastics from computers and televisions. Electronics are dismantled in nearby villages as part of a massive recycling industry outside Hebron in the Palestinian territories. The scale of this industry is enormous: roughly half of all the e-waste generated in Israel finds its way to a cluster of four villages in the area. About 80% of households there — including both adults and children — are involved either directly or indirectly in processing e-waste to extract copper and other valuable metals.

The informal, unregulated trade takes a heavy toll. Hundreds of e-waste burn sites are scattered about the region and have polluted the soil with lead, as well as dioxins and other toxic compounds. “The landscape is saturated with these contaminants,” says Yaakov Garb, an environmental scientist at Ben-Gurion University of the Negev in Sde Boker, Israel, who has spent the past five years mapping the burn sites and assessing their effects on the health of people who live nearby. “Most houses are within a stone’s throw of a current site or former site.”

Locals say that the burning sickens entire villages with respiratory illnesses, and water that runs off the contaminated hillsides kills vegetation. Although medical information there is spotty, preliminary studies have found very high lead levels in children. And studies of exposure to similar e-waste sites in China have documented a range of health effects, including increases in spontaneous abortions, still births, DNA damage and breathing difficulties¹.

Now, an innovative plan is taking shape to clean up the electronics recycling industry in the Hebron Hills. Garb is working with local leaders, government agencies and nongovernmental organizations to remediate the toxic sites and replace burning with non-polluting recycling methods that still allow residents to earn a living. The Swedish International Development Cooperation Agency (SIDA) plans to

Palestinians burn refrigerator motors to extract valuable metals in the village of Idhna. Locals blame health problems on e-waste recycling.

provide US\$2.7 million to support the project, which is awaiting approval by the Palestinian Authority. And in a rare example of cooperation, the Israeli and Palestinian governments are nearing agreements to put the plan into action, says Garb.

“What he’s been able to pull together is nothing short of miraculous; getting both sides to agree like that at all levels, it’s remarkable,” says Richard Fuller, chief executive of the non-governmental agency Pure Earth in New York City, which will be overseeing the project. The Palestinian clean-up scheme, he says, could serve as a model project for similar toxic e-waste sites in poor communities around the world, where millions of tonnes of electrical equipment gets dumped each year.

BORDER CROSSINGS

Garb has a long history of working across borders — both international and academic. Born in Johannesburg in 1960, he came to Israel as a 13 year old with his parents, who saw no future in apartheid-era South Africa. He has studied irrigation in sub-Saharan Africa, former toxic waste sites in the Czech Republic and deforestation in Guatemala.

E-waste caught his attention in 2008, when he noticed something odd while conducting a freight survey in the southern West Bank. Each morning, 70 to 80 trucks left the Hebron area and passed into Israel. They returned to the West Bank each night, laden with washing machines, refrigerators, toaster ovens, LCD screens, computers and furniture.

After publishing his traffic survey, Garb set the data aside for a few years, until he did a study analysing the water sources in hundreds of West Bank towns and villages. Residents of the town of Beit Awwa told a student working with Garb: “We used to collect the rainwater, but we



HEDI LEVINE FOR NATURE



“What he’s been able to pull together is nothing short of miraculous; getting both sides to agree like that at all levels, it’s remarkable.”

don’t because of the ‘black rain.’ Laundry, hung out to dry, came back soot-coloured. The villagers blamed the black rain on the local recyclers, who burn electrical cables and wiring to extract copper.

That made Garb even more curious. He and another graduate student, John-Michael Davis of Memorial University of Newfoundland in St John’s, Canada, analysed the local e-waste economy from top to bottom by conducting hundreds of interviews and doing randomized surveys. Davis even moved to Beit Awwa for more than a year to immerse himself in the community. Their study² is the first to chart the entire e-waste recycling economy anywhere, says Garb. They found that the items in the trucks came from a variety of sources in Israel: technology

companies or government ministries upgrading computers or routers and disposing of the old equipment; repair labs that discard or sell old televisions or other hard-to-fix items; and households getting rid of old appliances. Some of it ends up in Beit Awwa, says Davis, where there is a big market for discarded furniture and appliances. Although many items are refurbished and resold, large quantities of electronic waste are dismantled and burnt in an industry spread across Beit Awwa and three adjacent villages: Idhna, Al-Kum and Deir Sammit. He and Garb calculated that some 60,000 tonnes of e-waste — about half of what is produced in Israel — were processed at these four villages in 2014.

Some residents break down or repair the equipment in informal facilities, and others burn components at more than 500 sites in and around the villages. About 70 of these sites have, at some point in the past decade, burned at least a tonne of waste per day over the course of a year, says Garb. The local industry took off in 2004 after the price of copper jumped and Israel’s construction of a security barrier made it difficult for Palestinian men in this area to cross into Israel for work.

TOXIC IMPORTS

This kind of e-waste traffic from prosperous to poorer communities happens around the world. According to a report³ compiled by the United Nations, an estimated 41.8 million tonnes of waste electrical and electronic equipment was generated in 2014 (See ‘Mountains of

e-waste'). Europe and Asia are the largest producers, and African and Asian countries — including Ghana, Nigeria, China, Pakistan, India and Vietnam — are key destinations for shipments of hazardous e-waste for dumping.

In the Palestinian territories and much of the developing world, recyclers rely on inexpensive methods — using hammers and axes to dismantle equipment and burning cables to extract the copper. These techniques are also among the most polluting. So when Garb became aware of the problem in what is practically his backyard, he felt compelled to try clean it up, he says.

In combination with Palestinian and Israeli officials, Garb and his colleagues are attempting to transform the illegal, unregulated enterprise into a formal recycling trade — with facilities that allow for safe extraction of valuable components. Instead of stamping out the industry, which he says would drive it elsewhere in the West Bank, Garb hopes to build a partnership that will benefit all parties.

A tour of Beit Awwa and surrounding villages reveals how much is at stake. An acrid smell fills the air as Garb walks past ramshackle workshops and blackened hills. One facility is stacked high with old lamps, refrigerators, sinks, cables, metal cabinets, motherboards, keyboards and a pile of burned metal scrap. Nailed to a pole outside, an old Hebrew sign warns, “Please keep [area] clean.”

A boy, about ten years old, saunters about in flip-flops over the greasy black ground. According to a small survey by Garb, the average age at which burners start in the industry is 15.

Many residents worry about the recycling industry. At a burn site close to a quarry east of Idhna, two men driving past in their truck stop to talk to Garb. They complain that rainwater run-off from the fires has rendered nearby fields so contaminated that crops cannot grow there. During burning, chickens die or lay eggs without shells or yolks, say the workers.

In an unpublished study, Garb and Davis found hints that one type of cancer might be more common near burn sites. Garb has worked with a local group, Al Yassaria Women's Association, as well as Noam Weisbrod, a contaminant hydrologist from Ben-Gurion University, to submit a proposal to the US Agency for International Development to study household exposure to e-waste burning and its health effects, including birth defects and cancer.

Preliminary results from other studies suggest that children in the area are getting high doses of heavy-metal pollution. Sulaiman Swaiti, a Palestinian nurse from Beit Awwa who is now a master's student in public health and environmental studies at Ben-Gurion University, analysed levels of lead in the blood of 22 children from Deir Sammit. Twelve children had concentrations above 5 micrograms per decilitre, the point at which the US Centers for Disease Control and Prevention recommends public-health action; the highest value was 18.7 micrograms per

decilitre. For comparison, a separate study⁴ of children in the West Bank found an average lead level of 4.2 micrograms per decilitre.

Swaiti plans to do a more formal study by measuring lead in blood samples from 40 children in each of the other villages, plus a control group. He will also test dust in their households for heavy metals, and correlate the results with school grades.

“We have been suffering from this e-waste for a very long time — from the emissions from the burning of plastics, copper and aluminium,” says Muhamad Sweity, administrator of the Al Yassaria municipality, which covers Deir Sammit, Beit Awwa and Al-Kum. “We suffer from many, many problems: contamination of water, agriculture, farm animals, nature.” Children who live close to the burn sites have breathing problems, he says, and olive-tree yields have declined year after year. “We tried several times to solve the problem with the aid of the police, the [Israeli] Civil Administration, the government — nothing helped us,” Sweity says.

Although burning e-waste is illegal, oversight in the area is complicated because local Palestinian police have to coordinate with the Israeli Civil Administration — which has military and administrative control of the area — to enter the region where much of the burning takes place. By the time police get there, the people responsible are typically gone. And many families earn their livelihood from recycling, so they are reluctant to give it up. What is really needed, Sweity says, is financial support for clean recycling companies, as well as better police enforcement and inspection schemes.

TWO WORLDS

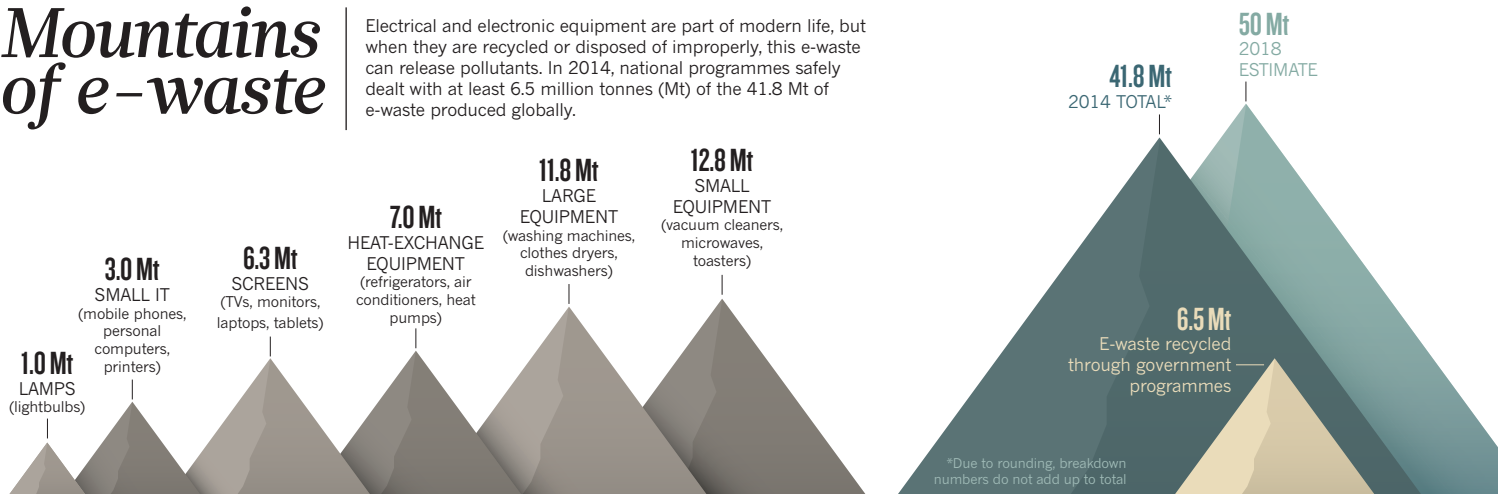
That's where Garb comes in. With his calm and easy-going manner, he gets a warm welcome in the Palestinian villages and also works effectively with Israeli authorities and other parties. “Somehow I seem to be put together in a way that allows me to move between the Arab and Jewish worlds, and between the social worlds of consulates and scrap yards, ministries and smugglers,” says Garb.

That sensibility has helped to give Garb a unique role in this area, says Johan Schaar, head of development cooperation at the Consulate General of Sweden in Jerusalem. Sweden has been active in promoting development in the Palestinian territories, and Schaar first met Garb three years ago. “What Yaakov has done, all this mapping that he has been able to do in these villages, that he has gained the confidence of all the people involved in this, is quite extraordinary,” he says.

Garb's efforts stand out, say others in the region, because there is a high degree of tension there, with little progress towards peace after nearly 50 years of Israeli occupation of the West Bank. In December 2015, SIDA gave Garb and Pure Earth a \$180,000 grant that funded a trial remediation of two sites in Beit Awwa. Workers used a tractor to scrape off the black surface goop, and then dug out the remaining toxic

Mountains of e-waste

Electrical and electronic equipment are part of modern life, but when they are recycled or disposed of improperly, this e-waste can release pollutants. In 2014, national programmes safely dealt with at least 6.5 million tonnes (Mt) of the 41.8 Mt of e-waste produced globally.





A young worker collects wiring from e-waste in Idhna.

“We suffer from many, many problems: contamination of water, agriculture, farm animals, nature.”

grime with picks and shovels. They transferred the material to a double-sealed plastic storage deposit at the same site. (Garb is now negotiating to move the contaminated soils to a certified facility for hazardous-waste disposal in Israel.)

That was just the first phase of a much broader project. In January this year, Garb and his team drew up an ambitious programme consisting of three components: clean up the hazardous-waste sites, create a sustainable Palestinian recycling sector and prevent further contamination of sites. The plan to build a sustainable recycling industry was key to convincing the Israeli Civil Administration to come on board, Garb says.

Schaar says that SIDA expects to provide a \$2.7-million grant, channelled through Pure Earth, to implement the broader plan. One part will be used to clean up 100 toxic waste sites. Another portion will fund an interim programme of free, legal copper recycling for residents of these villages. And a third will pay for a small rapid-response unit to shut down illegal burning.

Fuller says the project yields multiple benefits by “giving jobs to

the poorest, helping to build that economy, and that helps the whole peace process”.

The Israeli government and Palestinian Environmental Quality Authority are now in the final stages of negotiating an agreement to put the plans into place.

BACK TO THE GRIND

The cleaner future that Garb envisions is slowly taking shape in a small warehouse in Idhna. At the Safa Recycling and Material Processing company, workers shovel wiring from electronics and electrical cables into a \$220,000 grinding machine that shreds copper cables and separates them from their plastic sleeving. Safa's owner, Ismail Suleiman, says that when he bought the machine, “everyone was happy and clapping”, but it has been difficult to turn that into a profitable business. Grinding the cables costs much more than burning them. And he had trouble obtaining the proper permits to import the cables legally from Israel, he says. “There is no enforcement of laws by the Palestinian Authority against burning,” he says. “Any way you look at it, it's a mess.”

Money from the Swedish pilot project has helped. In February, the SIDA grant provided funds for locals to bring in cables for free grinding. Recyclers were so eager that they processed 15 tonnes in just 3 days, demonstrating that they might embrace a cleaner alternative to burning if it is economical. A portion of the new Swedish grant would provide an economic incentive to continue this legal recycling.

The Israeli government has also become involved in the clean-up efforts, because smoke from the four villages affects Israeli residents as well. Benny Elbaz, the Israeli Ministry of the Environment official who heads the West Bank environmental division of the Israeli Civil Administration in Beit El, is upbeat about the upcoming agreement. “We're doing everything we can to ensure the success of this project,” he says.

Elbaz plans to expand the initial pilot recycling project at Safa to other materials and other companies. Garb, too, has high hopes for extending the scope of the industry. He foresees “competitive boutique recycling”, in which Palestinians develop their own micro-niches for hand-dismantling devices and extracting rare-earth metals such as neodymium from the powerful magnets found in microphones and hard drives.

Pure Earth sees the Palestinian–Israeli initiative as a model for how to clean up burn sites in urban centres of many other poor places around the world, including those in Africa. Fuller says that Garb's project is one of the best designs he has seen, because it is politically viable. “I hope that we're going to do it in many other places,” he says.

But making the shift from burning to clean recycling is challenging, says Kees Baldé, an associate programme officer at the United Nations University in Bonn, Germany, and co-author of the 2014 report³ on global e-waste. The higher costs of clean recycling and local corruption in many areas often conspire to doom such efforts. If the Palestinian–Israeli agreement succeeds, Baldé says, “I think that the societal impact is going to be big”. With proper facilities, e-waste recyclers will earn more money, and be protected from the toxins released by burning, he says.

Even if it works well, the project in the Hebron Hills will not completely solve the problem there, says Garb, although it will put a big dent in it. The blackened slopes around Beit Awwa and nearby villages show just how far he and his colleagues have to go in cleaning up the pollution. But here and there, a delicate purple wildflower or a green shoot pokes up amid the scattered phone cases and burned rocks — a sign of what could be. And the agreement over e-waste is a rare case of Israelis and Palestinians working together for a better future, says Garb, “to see beyond the politics of the moment to the long-term human and environmental significance”. ■

Josie Glausiusz is a science journalist in Israel.

1. Grant, K. *et al.* *Lancet Glob. Health* **1**, e350–e361 (2013).
2. Davis, J. M. & Garb, Y. *Techniques Cult.* **65–66**, 390–403 (2016).
3. Baldé, C. P., Wang, F., Kuehr, R. & Huisman, J. *The Global E-waste Monitor — 2014* (United Nations University, IAS — SCYCLE, 2015).
4. Safi, J. *et al.* *Environ. Health Perspect.* **114**, 917–922 (2006).

COMMENT

DEVELOPMENT A research agenda for shifting the food system towards nutrition **p.30**



SUSTAINABILITY Current food metrics are reductive and distorting **p.33**

PUBLISHING How university presses are adapting to survive **p.35**

OBITUARY Susan Lindquist, protein folding visionary, remembered **p.40**



Hong Kong offers subsidies and reduced berthing fees for ships that switch to low-sulfur fuel in its harbour.

Four routes to better maritime governance

Challenges in preventing pollution from ballast water highlight reforms needed in global shipping regulation, write **Zheng Wan** and colleagues.

Sea freight carries more than 90% of global trade — and thousands of unwelcome passengers. The ballast water that stabilizes marine vessels is the greatest source of harmful bacteria and invasive species in aquatic ecosystems. About 10 billion tonnes of ballast are transported globally each year, with 7,000 species carried onboard every day¹.

This is damaging marine biodiversity and public health. For example, the imported

Chinese mitten crab (*Eriocheir sinensis*) is endangering native European and North American crayfish and salmon. Transported algae can seed blooms that smother or poison aquatic life, contaminate seafood and foul drinking water². As the world's shipping lanes expand into a warmer Arctic, invasive species will spread to waters that were previously unreachable³.

Managing ballast discharge requires

worldwide legislation and enforcement. International shipping traverses the high seas, where there is no local jurisdiction. A vessel registered in one country can operate thousands of kilometres away.

Global action has been slow. The 8 September accession of Finland to the International Convention for the Control and Management of Ships' Ballast Water and Sediments ended a 27-year slog to bring the treaty into force ▶

► (see ‘Too little, too late’). With 52 members of the International Maritime Organization (IMO) signed up — crossing the threshold of more than 35% of the world’s shipping tonnage — the treaty will finally come into force on 8 September 2017. After that date, any ship from a signatory state found to be violating the convention within regulated waters will be warned, fined or detained.

This month, the IMO council meets in London to coordinate and plan the organization’s activities. We argue that it must take a close look at the ballast-water convention, whose inadequacies highlight fundamental problems with international maritime governance. The lessons learned might steer other global environmental policies, from reductions in greenhouse-gas emissions to mitigating acoustic and light pollution. Going forward, the IMO should develop strategies to ensure that nations enter into its conventions promptly and to coordinate regional actions. It should establish market instruments to provide incentives and reform how maritime data are collected and used.

THE PROBLEM

The IMO’s ballast-water convention was shaped largely by polluters, not their victims. As of 1 January 2015, the top 35 ship-owning regions — including Greece, Japan and China — control almost 95% of the world’s shipping by tonnage. Yet more than 70% of their total capacity is registered under foreign flags, usually of nations with lower environmental standards (see ‘Tonnage top ten’).

Some of the first nations to sign up to the convention were the Maldives, Spain and

Nigeria, countries with strong interests in marine ecology but tiny fleets. Nations that are invested in the shipping industry have been lethargic. Panama, for example, which has more than 18% of the world’s registered tonnage, acceded to the convention only in October, owing to strong industry lobbying.

A few nations have made individual efforts to limit ballast-water damage. Since 2001, Australia has prohibited trading ships from dumping foreign ballast water within its coastal limits. Incoming vessels must flush their tanks in the open ocean, and report and verify that they have done so. Brazil, Canada and Israel have similar measures⁴.

The United States has gone further. It requires that any foreign ballast water dumped in its waters by commercial vessels must meet the IMO standards, but coastal states may set controls, too⁵. California has proposed the most stringent limit: no detectable living organisms more than 50 micrometres across. Because this is impossible to meet with current treatment technologies, this standard has been postponed until 2020.

Such uncertainties and more rigorous testing by the US Coast Guard are slowing down the approval of ballast-water treatment systems. As of the most recent listing in May 2016, none of the 65 designs approved by the IMO had received US approval. Until some do, ships with IMO-approved systems may not comply with US standards. As an interim measure, by July, the US Coast Guard had

“The lessons learned might steer other global environmental policies.”

listed 56 IMO-approved systems that can be used in its waters for a 5-year grace period.

Meanwhile, the owners of 70,000 ships globally do not know what to invest in. Ballast-water treatment equipment is expensive. It involves physical and chemical processes such as filters, chlorine, ozone or ultraviolet light. A unit can cost US\$1 million, with adaptation and maintenance costs on top. Early adapters will be penalized because they will have to pass on those costs to cargo owners.

Naturally, the maritime industry prefers to wait and see. There is no economic incentive to do otherwise. Plus, many countries, especially developing ones, lack the technology to test the quality of ballast water and the capacity to punish those who flout the law.

Risk assessments that would allow targeted solutions are missing. The consequences of invasive marine species vary between regions, and depend on traffic volume and ships’ ports of call⁶. Data are scant. We need to know which species live in particular ports or waters, how temperature and salinity influence species’ survival elsewhere and where and how often vessels move.

Emerging technologies could help, but are limited. For example, the Automatic Identification System publicly transmits a ship’s identity and position in real time to avoid collisions at sea⁷. But not all ships are equipped, and the system is prone to human error. Historical data are expensive for researchers to acquire because they are provided by commercial vendors, and mainly to big shipping companies.

THE FIX

Implementing the following four reforms would improve global maritime governance and marine biodiversity conservation.

Speed entry into force. The IMO should activate future conventions faster and more fairly. Lowering the minimum registered tonnage limit would bring forward implementation⁸. For example, had the ballast-water convention required coverage of at least 30%, rather than 35%, of the world’s shipping tonnage, it would have been triggered in 2013 by Germany’s accession (the 37th signatory). Increasing the minimum number of states to sign up at the same time would also rebalance voting weights and raise the priority of environmental protection over fleet size.

Coordinate regional actions. The IMO should harmonize standards at least regionally. It should provide technical assistance and guidance so that coordinated actions are constructive, coherent and well-administered. Regional actions can have a wider influence and encourage more data-driven and evidence-based policies. For example, European Union regulations adopted in 2015 that set out a legal framework for monitoring,

TOO LITTLE, TOO LATE

A decades-long crawl to global consensus

In 1969, the World Health Organization (WHO) adopted rules for trade and travel to prevent the global spread of epidemics such as plague and cholera through ships and aircraft.

In 1990, Canada and Australia put the issue of ballast water on the agenda of the International Maritime Organization (IMO) when they became concerned about invasions by Eurasian zebra mussels (*Dreissena polymorpha*) and Japanese dinoflagellates. The IMO issued a voluntary protocol suggesting that ships flush their ballast tanks 200 nautical miles (370 kilometres) from land. Few ship owners took notice. Following the 1992 United Nations Earth Summit and the adoption of the Convention on Biological Diversity, the IMO began to develop an internationally binding instrument.

In 2004, the IMO adopted the

International Convention for the Control and Management of Ships’ Ballast Water and Sediments. It includes a roadmap for signatories to phase out existing methods (such as tank flushing) in favour of stringent standards of water treatment — an upper limit to the permitted concentration of organisms that can remain in treated ballast water depending on ship type.

For example: ships should discharge fewer than 10 viable organisms per cubic metre for organisms measuring 50 micrometres or larger, or 10 organisms per millilitre for those 10–50 micrometres in size. Limits are also set on the numbers of specific microbes, such as *Escherichia coli* and intestinal enterococci.

Ship owners must invest in water-cleansing equipment that filters large particles and kills organisms and pathogens by exposing them to ultraviolet light, ozone or chlorine.

TONNAGE TOP TEN

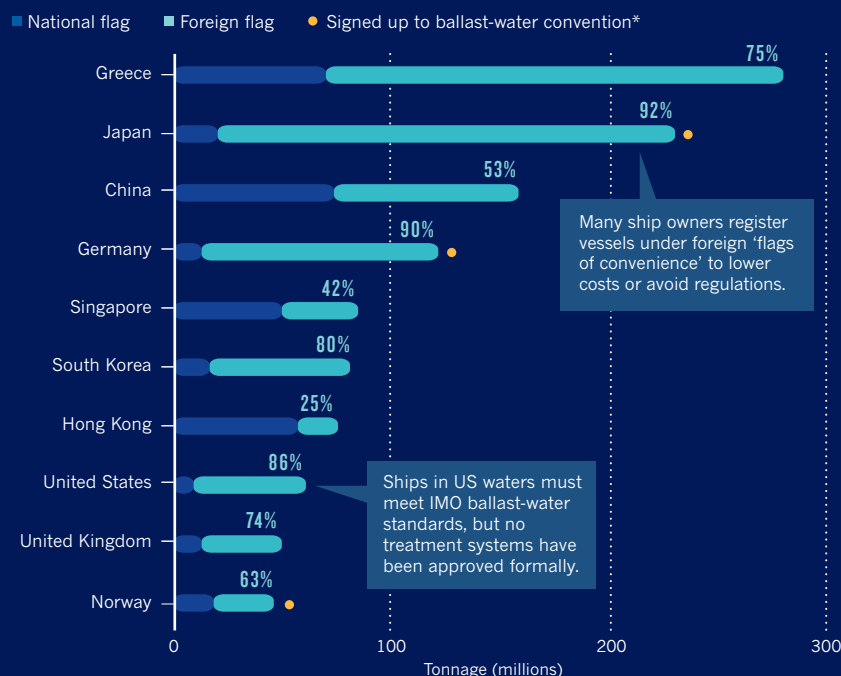
Ship owners from ten regions control more than two-thirds of the weight carried by the world's fleets. Two-fifths of this tonnage is ballast water, which contains marine organisms that are spread when discharged. Only three of these major shipping areas have so far* acceded to the 2004 International Maritime Organization (IMO) ballast-water management convention, which enters into force in 2017.

Global ship ownership and carrying weights

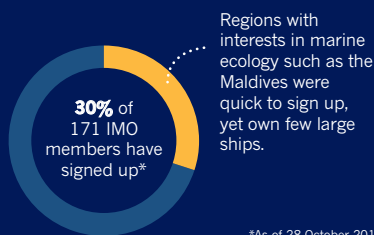
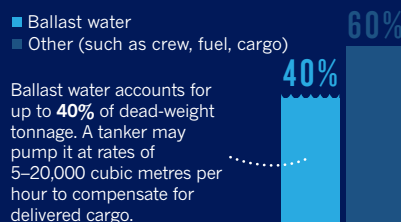
Owners in 10 regions account for **68%** of the world's dead-weight tonnage, which includes cargo, fuel, fresh and ballast water, provisions, passengers and crew.



Areas with most dead-weight tonnage



Managing ballast water and the environment



*As of 28 October 2016

reporting and verifying greenhouse-gas emissions from shipping have inspired the IMO to develop a similar policy.

Develop market instruments. Today's IMO conventions for pollution prevention and control are technology-oriented — they stipulate standards, but say little about how to fund or achieve those measures. Industries need a mixture of technical standards and market instruments to provoke them to transform⁹. Incentives could include reduced

port-usage fees, priority access to port facilities and tax rebates for early adapters. For example, more than 12% of the vessels visiting Hong Kong since 2012 have switched to low-sulfur fuel while berthed, in exchange for lower port fees and subsidies. Developed countries should provide technical and funding aid for developing countries.

Reform data usage. The IMO and regional maritime bureaucracies must improve how data are collected, shared and used.

Developments in ship tracking, global-positioning sensors, the Internet of Things and 'big data' should be incorporated into models that predict maritime pollution. By analogy, Singapore's road-pricing scheme uses devices to detect cars and to debit fees, as well as big-data analytics to calculate the optimal toll for controlling traffic.

Real-time data analyses would help ship operators and authorities to detect malfunctions in pollutant treatment systems (such as for sewage, garbage or gas emissions) more quickly, without inspections or business disruptions. The shipping industry will benefit from more-efficient operations and smarter decision-making on energy saving and pollution control.

Governmental funding will be needed to set up platforms and protocols for data collection and sharing. But investments will be offset by savings from avoiding self-reporting, which is labour-intensive, burdensome and unreliable.

As the pace of global change and trade accelerates, we cannot afford to wait decades to negotiate how to deal with each type of maritime pollution. ■

Zheng Wan is associate professor at the College of Transport and Communications, Shanghai Maritime University, Shanghai, China. **Jihong Chen** is professor and maritime expert at the Science and Technology Committee of Shanghai Municipal Transport Commission, Shanghai, China. **Abdel El Makhoulfi** is senior economist at the Amsterdam University of Applied Sciences, Amsterdam, the Netherlands. **Daniel Sperling** is distinguished professor of civil engineering and environmental science and policy, and founding director of the Institute of Transportation Studies at the University of California, Davis, USA. **Yang Chen** is associate professor at Shanghai International Shipping Institute, Shanghai, China.
e-mail: mrwan@ucdavis.edu

1. Tamelander, J., Riddering, L., Haag, F., Matheickal, J., & No, G. M. S. *Guidelines for Development of a National Ballast Water Management Strategy* (GloBallast Partnerships & IUCN, 2010). available at <http://go.nature.com/2fftcqu>
2. Smayda, T. J. *Harmful Algae* **6**, 601–622 (2007).
3. Miller, A. W. & Ruiz, G. M. *Nature Clim. Change* **4**, 413–416 (2014).
4. Lloyd's Register *National Ballast Water Management Requirements* (Lloyd's Register, 2016).
5. Albert, R. J., Lishman, J. M. & Saxena, J. R. *Ecol. Appl.* **23**, 289–300 (2013).
6. Xu, J. et al. *Proc. 20th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* 1699–1708 (2014).
7. McCauley, D. J. et al. *Science* **351**, 1148–1150 (2016).
8. Tan, A. K.-J. *Vessel-Source Marine Pollution* (Cambridge Univ. Press, 2005).
9. Wan, Z., Zhu, M., Chen, S. & Sperling, D. *Nature* **530**, 275–277 (2016).



EDUARDO SOTERAS/AFP/GETTY

A woman feeds her malnourished child in the Democratic Republic of the Congo.

A new global research agenda for food

Lawrence Haddad, Corinna Hawkes and colleagues propose ten ways to shift the focus from feeding people to nourishing them.

Around 57 of the 129 countries that have data on undernutrition and obesity are struggling with both¹. Everywhere, the consumption of vegetables, legumes, fish, nuts, seeds and fruits is much below that recommended by the World Health Organization (WHO). Meanwhile, people are consuming too much fat, processed meat, salt and sugary drinks.

Global food systems are failing to keep us all fed, let alone healthy. How food is grown, distributed, processed, marketed and sold determines which foods are available, affordable and desirable. These factors have a crucial role in the quality of people's diets, and hence play a vital part in health.

Diet is the number one risk factor in the global burden of disease². Poor diets are

responsible for more of the global burden of ill health than sex, drugs, alcohol and tobacco combined. In the next few decades, food systems will be under further stresses from population and income growth, urbanization, globalization, climate change and increasingly scarce natural resources.

Although 795 million people are undernourished and lack essential vitamins and minerals³, obesity is behind many of the chronic diseases that are sweeping the globe, from type 2 diabetes to heart disease. One in three people is malnourished. Almost one-quarter of children under five have stunted growth, with diminished physical and cognitive capacities. Across Africa and Asia, the impact of undernutrition on gross domestic product is 11% annually¹. At the same time,

2 billion adults worldwide — more than 1 in 4 — are overweight or obese.

This is not a problem that countries can overcome through growth or development. As economies expand, many social factors improve, but the quality of diets does not. Hunger and famine have fallen substantially, thanks to rapid poverty reduction and rising agricultural productivity. But progress remains too slow in many respects. Moreover, middle- and low-income countries are now following the well-worn, highly damaging path from undernutrition to obesity.

The efforts required from the international community are equivalent to those marshalled to tackle HIV/AIDS, malaria and smoking. In particular, urgent interdisciplinary research is needed to support concerted

policy action. That is our conclusion from compiling a report commissioned by the Global Panel on Agriculture and Food Systems for Nutrition and published in September⁴. It includes a call to scientists, governments and donors to work out how to craft and sustain food systems to provide nutritious diets for all.

As the United Nations Food and Agriculture Organization (FAO) and the WHO meet to discuss the UN's Decade of Action on Nutrition in Rome this week, researchers, governments, industry experts and funders must commit to meeting these challenges — which are inextricably knitted with the Sustainable Development Goals, especially goal 2 (zero hunger) and goal 3 (good health and well-being). Piecemeal action will not do: the trends are so large and interconnected that the entire food system needs overhauling. No single government ministry or sector 'owns' the quality of diets available to the consumer. The lack of well-defined public-sector responsibility in many countries is all too clear.

TEN RESEARCH PRIORITIES

Here we set out a new global research agenda for nutrition. It is aimed mainly at researchers, funders and governments, but has important messages for all stakeholders.

Identify entry points for change. Food systems are dynamic; they comprise production, storage, transport and trade, processing, transformation and retail and the interactions between these. We need to understand how each part of the system can contribute in an integrated way to making high-quality diets more available, affordable and appealing. For example, how can we reduce food spoilage in storage, minimize nutrient losses during transformation and improve food quality with minimal processing? Studies should involve multi-country and site-specific appraisals, qualitative research and mapping, and thorough analyses of value chains. Donors and funders should promote initiatives that are interdisciplinary, involve consumers and policymakers, and analyse which incentives cause actors in the food system to behave differently.

Make more data on diets widely available. It is currently difficult to compare diets across cultures, geographies and time. This has hampered a global consensus on what constitutes a healthy diet. A pilot project — the FAO/WHO Global Individual Food consumption data Tool (FAO/WHO GIFT; see go.nature.com/faogift) — aims to answer some elements, but has too few resources to be truly effective. The project needs a larger team to collate many more national surveys and develop guidelines for future surveys. Other initiatives include the Global Dietary Database project and a proposed



Farming on an industrial scale uses feedlots such as these near Yuma, Arizona.

World Gallup Poll initiative to include questions on diet quality. Such efforts are poorly funded and piecemeal. What is needed are open-access portals for diet data, such as the WHO's global database on child malnutrition. This incorporates hundreds of surveys of child heights and weights, and has proved invaluable for documenting progress and holding stakeholders accountable.

Agree on what constitutes a healthy diet.

People do not choose nutrients, they select combinations of foods in differing amounts. Pairings of single foods and diseases are the basis of risk-factor analysis in global burden

"It is difficult to compare diets across cultures, geographies and time."

studies, but tell us little about diets as a whole. Although there are studies on the value of, say, the Mediterranean diet, there are few from low-income countries. And even the nutritional profile of many important indigenous foods remains poorly known. A better understanding of dose-response relationships is needed. Is it better to eat a little of each food category frequently, or a large amount less often? Food researchers need to be more creative and research funders bolder in assessing the health implications of common combinations of foods.

Tackle different forms of malnutrition simultaneously. Credible information on what works to reduce both undernutrition and obesity (and related diseases) is essential if governments and industry are to scale up investments aimed at improving diets. Successful programmes that tackle undernutrition tend to be on a small scale⁵. Evidence on how to address unhealthy diets

associated with obesity are often population-based and thus harder to evaluate. There is analysis on the effect of interventions to improve the availability, affordability and appeal of food, such as taxes, school-food standards and nutrition labelling⁶. But the impact of these on population subgroups or on obesity is little studied. The most serious gap concerns interventions farther back in the food system, such as standards and new incentives for more-effective transport and retail of nutritious foods (see also 'Identify entry points for change').

Understand the role of chain length. In 'short-chain' models, food passes directly from those who grow or rear it to those who eat it, such as in subsistence-farming contexts, at farmers' markets, or where school-meals programmes source food from family farms. In 'long-chain' systems, foods travel great distances or go through multiple transactions to reach consumers. What combination of these systems ensures high-quality diets — enough of the right kind of food at the right price, for all? Research must probe sustainability concerns, such as how to reduce food waste and greenhouse-gas emissions, and investigate the social and economic benefits of different types of system (see also 'Fix metrics').

Analyse business incentives. Apart from consumers, the most numerous stakeholders in global food systems are private entities — from farmers to food processors, wholesalers and retailers. In 2015, just 4 of the top 25 food and drinks firms made one-third of the total sales. Thus, the private sector could help to tilt food systems towards higher-quality diets, and could respond innovatively to targets and regulations. We need to understand what incentives will tip these levers in a healthier

► direction. We need better mechanisms for public–private dialogue to shape and implement research priorities. Collaborations between competitors during the early development of commercial products could shift sectors collectively and overcome disincentives for single firms to act in isolation. One such example hosted by the Global Alliance for Improved Nutrition is a programme called Business Platform for Nutritious Research, which seeks to substantially increase business investments in nutritious products and services by addressing evidence gaps that limit companies' willingness and ability to make such investments. To strengthen accountability, research is needed on which public–private partnerships improve diets most. Such research exists in the health field⁷.

Account for climate. Global warming will affect the availability of different foods from land and sea, and their nutritional value. Diet changes will also drive significant shifts in land and water use and in greenhouse-gas emissions⁸. How might diversification of local food systems address agricultural resilience and nutritional diversity? The consumption of some of the most micronutrient-dense foods — including fruits, vegetables and animal proteins — must be increased in poor communities, but such foods place substantial demands on environmental resources. Research on less-polluting production of these, in the context of rapidly changing livestock systems, is crucial. Research could also identify ways of achieving big wins for environments and diets. Evidence suggests, for instance, that the adoption of WHO guidance on healthy diets could reduce global mortality by 6–10% and food-related greenhouse-gas emissions by 29–70% compared with a reference scenario in 2050 (ref. 9).

Study supply and demand. The availability of food is influenced by consumer preferences, relative prices and supply. Processed foods and sugar-laden beverages are found in remote areas of Nepal and Ethiopia; a choice of vegetables, fruits and fish is not. In part, this is because investment in improving the production and availability of agricultural commodities has for decades been focused on a small number of cereals. Funding in global public-sector research institutions is still focused mostly on rice, wheat, maize (corn) and other grains. About 45% of private-sector agricultural research investment is on maize¹⁰. Public and private research on neglected nutritious commodities — including fruits and vegetables such as mangoes, carrots and spinach, pulses such as lentils, fish products and seeds and nuts — needs to increase with a focus on their yield and resilience to pests, diseases and climate change. This is an important message for the global research community, broadly led

by the Consultative Group on International Agricultural Research (CGIAR). Donors should support the CGIAR's 2014 commitment to mainstreaming nutrition in all crop-breeding programmes, and its attempts to direct more research to healthy agriculture and food systems. Promoting supply has to be coupled with promoting demand. More research is needed to educate, inform and encourage consumers to make positive choices for healthful diets — in low- and high-income settings.

Identify the economic levers for change.

Every US\$1 spent on successful nutrition programmes offers roughly \$16 of benefits¹. But we do not know enough about where in the food system we should invest in any one policy, regulation or programme to generate

“More research is needed to encourage consumers to make positive choices for healthful diets.”

the largest net pay-offs. Policymakers in many countries, including Indonesia, Rwanda and Peru, are calling for such evidence. The economic gains from improving the food system and diet quality must be calculated for different sectors and sub-systems under various scenarios — increased demand, climate change, industry innovation and shifts in consumer preferences.

Fix metrics. Raising diet quality through food systems cannot be at the expense of other Sustainable Development Goals. Calls to price the ‘real’ cost of water consumption or carbon production have promoted understanding of the economic externalities of individual and government choices. We need the same approach for diets (see Comment, page 33). We must be able to answer questions such as how the overall benefits and costs of consuming locally sourced fruits and vegetables compares with importing them. The relationship between health, energy use, water use and greenhouse-gas emissions involves trade-offs, but the variety of foods in each category means there is great scope for finding wins — if fine-grained analyses are done.

We urgently need effective methods for measuring both the sustainability and nutritional value of diets. Current work on the carbon footprints of commodities should be extended to analyses of the whole food system. Also essential is transparent annual accounting of what countries are doing to improve different points in the food system and with what effect. There is a real opportunity for the FAO, African, Asian and Latin American development banks and the World Bank to develop scorecards to track how nutrition-sensitive national urban and rural food systems are.

WHAT NOW?

The era of commodity research aimed at feeding a starving world is over. A new era has begun that requires us to nourish everyone in ways that can be sustained environmentally, economically and culturally. Policymakers urgently need to recognize that diets are compromising economic productivity and well-being as never before. Delegates to the upcoming G20 and G7 meetings in 2017 should take collective responsibility for fixing our failing food system.

Funders who support agriculture and nutrition research must focus much more of their resources accordingly, doubling their current allocations to more-nutritious food systems by 2020. Scholars and journals must become more pluralistic in the methods and approaches that they support. We can only fix problems in our food systems if we diagnose them correctly. If we do not, the world's future health and economic problems will be very much greater than they are today. ■

Lawrence Haddad is executive director of the Global Alliance for Improved Nutrition, Geneva, Switzerland. **Corinna Hawkes** is director of the Centre for Food Policy, City University of London, UK. **Patrick Webb** is professor of nutrition at the Friedman School of Nutrition Science and Policy at Tufts University, Boston, Massachusetts, USA. **Sandy Thomas** is director of the Global Panel on Agriculture and Food Systems for Nutrition, London; and professor of science policy at the Science Policy Research Unit, University of Sussex, UK. **John Beddington** is professor of natural resource management at the University of Oxford, UK, and co-chair of the Global Panel on Agriculture and Food Systems for Nutrition. **Jeff Waage** is technical adviser and **Derek Flynn** is policy adviser to the Global Panel on Agriculture and Food Systems for Nutrition, London, UK.

e-mail: sandy.thomas@glopan.org

1. International Food Policy Research Institute. *Global Nutrition Report 2016: From Promise to Impact: Ending Malnutrition by 2030* (IFPRI, 2016); available at <http://go.nature.com/2g3er5u>
2. GBD 2013 Risk Factors Collaborators. *Lancet* **386**, 2287–2323 (2015).
3. FAO, IFAD and WFP. *The State of Food Insecurity in the World 2015* (FAO, 2015).
4. Global Panel on Agriculture and Food Systems for Nutrition. *Food Systems and Diets: Facing the Challenges of the 21st Century* (Global Panel, 2016); available at www.glopan.org/foresight
5. Bhutta, Z. A. *et al.* *Lancet* **382**, 452–477 (2013).
6. Hawkes, C. *et al.* *Lancet* **385**, 2410–2421 (2015).
7. Roehrich, J. K., Lewis, M. A. & George, G. *Soc. Sci. Med.* **113**, 110–119 (2014).
8. Tilman, D. & Clark, M. *Nature* **515**, 518–522 (2014).
9. Springmann, M., Godfray, H. C. J., Rayner, M. & Scarborough, P. *Proc. Natl Acad. Sci. USA* **113**, 4146–4151 (2016).
10. Fischer, T., Byerlee, D. & Edmeades, G. *Crop Yields and Global Food Security: Will Yield Increase Continue to Feed The World?* (ACIAR, 2014); available at <http://go.nature.com/2gkkez>



Drying red chillies under the sun provides one of the few sources of employment for women in an area of Bangladesh.

Fix food metrics

For sustainable, equitable nutrition we must count the true global costs and benefits of food production, urge **Pavan Sukhdev, Peter May and Alexander Müller.**

Current patterns of crop and livestock production and of processing, transport and consumption, are not delivering healthy, nutritious food to society. They are generating large and unacceptable impacts on the environment and on vulnerable populations.

Food systems are now the source of 60% of terrestrial biodiversity loss, 24% of greenhouse-gas emissions, 33% of soil degradation and 61% of the depletion of commercial fish stocks¹. And the increasing homogenization of food sources worldwide is narrowing the genetic diversity in animals and plants that is crucial to secure human nutritional needs against climatic and other changes.

Food systems are undermining human health all over the world, by permitting, or even promoting, inappropriate diets or unsafe foods. As the *Global Nutrition Report*, released this September puts it: “Diet is now the number-one risk factor for the global burden of disease”².

Around 800 million people in developing countries consume less than the 2,100 kilocalories per day recommended by the World Food Programme³ because of failures in access and distribution. At the same time, 1.9 billion people in the developed world take in more than 3,000 kilocalories per day⁴. As processed foods high in fat and carbohydrates come to dominate, even in developing nations, “the number of children under 5 who are overweight is approaching the number who suffer from wasting”².

Yet agriculture employs around 1.3 billion people. Around 1 billion work on small farms⁵ (less than 2 hectares). Admittedly, their working conditions could mostly not be rated as ‘decent’ by the International Labor Organization’s definition, but what are the chances that society could provide alternative employment to these people if their 500 million small farms gave way to concentrated, large-scale and highly mechanized agribusiness? The world is already short of about 200

million jobs⁶, and major industries such as steel and car making employ only 6 million and 9 million people worldwide, respectively.

Small-scale agriculture provides subsistence, employment and most of the food directly consumed by urban residents throughout the developing world. It also ensures that rural landscapes are conserved as a touchstone for cultural identity.

Current metrics for agricultural performance do not recognize or account for any of these costs or benefits. The emphasis on yields or profits per hectare is as reductive and distorting as is gross domestic product, with its disregard for social and natural capital. Food metrics must be urgently overhauled or the United Nations’ Sustainable Development Goals will never be achieved.

HOLISTIC EVALUATION

We contend that a sustainable food system has three key attributes. It should deliver adequate nutrition and health across all

levels of income and societal development. It should avoid significant negative ecological and environmental impacts. And it should ensure equitable access to land, water, inputs (such as seeds and fertilizer) and technical and financial assistance for the roughly 1 billion people who still depend on small farms for their livelihoods.

To achieve this, policymakers, researchers and citizens need more reliable and integrated information on the hidden costs and benefits — the ‘externalities’ — of the whole agri-food system, not just parts of it. A new project that aims to deliver such information was launched in October 2016 by The Economics of Ecosystems and Biodiversity (TEEB), a network of organizations that we are involved with. Called TEEB for Agriculture and Food (TEEBAgriFood), the project will attempt to provide a framework for a holistic evaluation.

Take, for example, maize (corn). The trend is to convert much of what is (over-) produced into starch and sugar. In conventional agricultural analysis, the improvements in yield per hectare per year in intensive maize-production systems are usually presented as the main indicator of success. More maize for fewer dollars up-front is also considered an important contribution to food security.

Of course it isn't that simple. TEEB-AgriFood has analysed three different maize-production systems (intensive industrial, organic and traditional), looked at how the product of these different systems is used, and charted their impacts on ecosystems, biodiversity and human health. For example, increased production of maize in industrial systems has a clear link (through lower prices for high-fructose corn syrup) to the increased use of maize-based sweeteners in soft drinks, a significant driver of obesity and diabetes. According to the World Health Organization, the global costs of type 2 diabetes are more than US\$850 billion per year.

To evaluate maize holistically, we should account for its market price, its contribution to health costs, the environmental impacts of increased greenhouse-gas emissions from mechanized farming and transport, and the impacts of fertilizer use on aquatic life. TEEB-AgriFood hopes eventually to make all these links visible.

POLICY PROBLEMS

We have conducted other initial evaluations and policy analyses for agricultural systems producing rice, palm oil, livestock (including beef and dairy cattle, and also meat from chicken and pigs), inland fisheries and agroforestry systems. In each case, we have found the same sorts of policy failure, and problems with food security and nutritional quality. These problems cut across current commodity-production systems

at different scales — from smallholder to corporate farming — and along different commodity chains.

For example, significant costs are associated with the expansion of oil-palm plantations into tropical rainforests in Indonesia. This sometimes entails the draining of peatlands, which can then catch fire. The resulting health risks from air pollution are severe, especially for children and older people. According to the World Bank⁷, the disruption to economic activity in 2015 alone cost the Indonesian economy an estimated \$16 billion — more than the annual country-wide value added by palm oil.

Similarly, the externalities of soya bean production in tropical rainforests in the Brazilian Amazon were so damaging that they led to a boycott and the development of a certification scheme. In 2006, the Brazilian soya bean crushers' association decided to embargo purchases of soya beans from recently deforested areas; international and domestic non-governmental organizations served as watchdogs to the spread of clearances. Supportive purchase policies by traders in Europe stimulated buyers to accept only rainforest-certified soya beans. Deforestation directly related to soya bean production has all but ceased (see go.nature.com/2gdrcqw).

We have found that there is a great deal of knowledge about the ways in which foods are produced and the inputs they demand. There is much less information about the impacts of these value chains on the environment and on social groups. Furthermore, our studies have generally not touched on what happens to these products once they leave the farm gate. TEEB-AgriFood and others need to probe the relative impacts of different transport and processing pathways, and how options for food disposal at each stage of the supply chain and after consumption affect the human footprint on the planet. And more work is needed on the health risks of components of these food systems.

Other pressing questions need to be addressed by the TEEB-AgriFood study and similar explorations. How can we escape our massive dependence on fossil fuels to produce nitrogen fertilizer and how could we enhance the natural nitrogen cycle? How could cattle feed be altered to reduce methane production from livestock without putting undue stress on land needed for food crops? How can we increase fish protein in the diet without exhausting ocean stocks and endangering inland aquatic

ecosystems? These and many other related questions must be looked at in a systemic, integrated way.

WHAT NEXT?

These information gaps need plugging if we are to make the transition to sustainable agri-food systems that nourish, provide energy, damage neither health nor environment, and support equitable access to resources. Extending the initial TEEB-AgriFood evaluations will require the synthesis of a vast body of experience and learning from numerous practitioners in policy, farming and agri-business. We hope to engage this wider community starting now through a call for interest to authors over the next two years (see go.nature.com/2fwrqih).

To catalyse further research by others we will publish a series of studies that aim to explain and quantify the externalities associated with food systems and what can be done to reform them. To shift perspectives among consumers, business leaders and policymakers, we plan a raft of engagement exercises.

We call on researchers and practitioners who are as concerned as we are about the failings of today's agri-food systems to help us evaluate the alternatives. Our common goal should be to deliver safe and nutritious food to all, while respecting the communities and the fragile lands that nurture much of what we eat. ■

Pavan Sukhdev is a Goodwill Ambassador for the United Nations Environment Programme, founding trustee of GIST India, and former study leader of The Economics of Ecosystems and Biodiversity (TEEB). **Peter May** is professor of development, agriculture and society at the Federal Rural University of Rio de Janeiro, Brazil, a past president of the International Society for Ecological Economics, and a member of the Project Steering Committee for TEEB-AgriFood. **Alexander Müller** is a former assistant director-general for agriculture at the Food and Agriculture Organization of the United Nations, study leader of the TEEB-AgriFood Project, and managing director of TMG - Töpfer, Müller, Gassner, a sustainability think tank. e-mail: pavan@gistadvisory.com

1. Food Systems and Natural Resources (UNEP, 2016).
2. Global Nutrition Report: From Promise to Impact (Int. Food Policy Res. Inst., 2016).
3. The State of Food Insecurity in the World 2015 (FAO, 2015).
4. Alexandratos, N. & Bruinsma, J. 'World Agriculture Towards 2030/2050: the 2012 revision' ESA Working Paper No. 12-03 (FAO, 2012).
5. Lowder, S. K., Skoet, J. & Singh, S. 'What do we really know about the number and distribution of farms and family farms worldwide?' ESA Working Paper No. 14-02 (FAO, 2014).
6. International Labour Organization World Employment Social Outlook: Trends 2015 (International Labour Office, 2015).
7. The Cost of Fire: An Economic Analysis of Indonesia's 2015 Fire Crisis (World Bank, 2016).



The Barker Library at the Massachusetts Institute of Technology in Cambridge, seen in the 1930s. Changing habits are affecting publishers.

BOOK PUBLISHING

University presses adapt

Roger Schonfeld analyses how the sector's scientific books are faring in the digitized, open-access era.

University presses, those venerated producers of books (and journals), face pressures as never before. Digital access and changing acquisition patterns at libraries have disrupted the presses' traditional businesses, and they are meeting the challenges in different ways. Some are experimenting, some seeking shelter by joining with their academic libraries, and some maintaining impressive resilience. Others struggle to survive. Meanwhile, some presses retain healthy print-driven scientific books businesses, and are content to maintain them — as long as print books thrive.

The presses of the UK universities of Cambridge and Oxford, established in the sixteenth century, are respectively the world's oldest and largest university publishing houses. The dozens of US members of the Association of American University Presses are comparatively young — Johns

Hopkins University Press in Baltimore, Maryland, is 138 years old and the MIT Press in Cambridge, Massachusetts, 84. But their immense reputations are no guarantee against upset. Recent years have brought reorganization and lay-offs at Cambridge and Oxford, and several US institutions (including the University of Missouri in Columbia) have tried to shut their presses.

In the face of disruptions to and consolidation of scientific publishing in recent decades, some presses have carved out an important space in scientific book publishing. They have fostered key lists and promising authors in the natural and social sciences for a focused but not exclusively academic audience — unlike the big commercial publishers of popular science books, such as Penguin Random House in New York City, which concentrate on the mass market. And books remain among the most important

translational tools for science. They bring a scholarly examination to a public eager to know more about 'hot-button' research areas such as climate change, the potential of exoplanets, socio-economic inequality, artificial intelligence and genetics.

Academic book publishing rests on three pillars: the traditional formats of textbooks, monographs and general-interest titles. The latter are aimed at natural and social scientists in other fields, graduate students and lay readers. All are undergoing change.

With their high production values, prices and profits, undergraduate textbooks are still big business for firms such as Pearson in Cambridge, UK. For more than a decade, many have been more than just a book: they include access to online environments offering instructor notes, quizzes, tests and more. As publishers add multimedia access, interactive instructional modules and ▶

MIT MUSEUM

▶ other bolt-ons, this demands a new level of capital investment.

Universities and funders are mounting renewed efforts to compete with these businesses through open educational resources such as Open SUNY Textbooks at the State University of New York. Increasingly, textbooks follow a different trajectory from monographs and trade titles; their future is tied up with the delivery of digital instruction.

Meanwhile, monographs — relatively brief, expensive, sparsely edited and simply designed formats for original research — are aimed mainly at small peer readerships, often to enhance the author's chances of tenure or promotion. For stand-alone titles, sales (targeted almost exclusively at institutions) rarely exceed 500 and have declined as cash-strapped libraries shift resources. So scientific monographs are increasingly sold in series, such as the Princeton Series in Modern Observational Astronomy at Princeton University Press in New Jersey, or in partnership with a learned society — Cambridge University Press's Econometric Society Monographs, for example. The 'series' strategy enhances specific fields of strength; by putting 'more wood behind fewer arrows', the presses can leverage editorial expertise, author lists and reviewer relationships.

The third pillar — the general-interest, or academic trade book — is a case apart. These are "written with ambition to impress a wider readership beyond the tenure committee", explains Chris Harrison, publishing director at Cambridge University Press. Their authors, he adds, may be researchers "frustrated by the limits of the journal article as a format for communication who now want to join the dots and tell a bigger story".

These books translate a concept, idea or methodology and open up the scholarly literature. Notes Amy Brand, director of MIT Press: "There is a sweet spot for academic publishers in the trade space — capturing and valuing the complexity of the work yet not having to write for an eight-year-old, respecting the intelligence of the general reader."

TRADING PLACES

Some publishers may view these as works for the generalist, but Peter Dougherty, director of Princeton University Press, sees a more specific objective. "We are trying to reach a broad cross-section of scholars, beginning with those in the author's own field and moving into contiguous fields." These would once have been published by trade houses, but as commercial publishers have focused on higher-selling titles, academic trade books have become an opportunity for university presses. A small group of presses, including Cambridge and Princeton, have developed a substantial business in these books, but they can demand a lot from the publisher.

The authors with the most well-honed



Publishers strive for a share of the market at the London Book Fair.

concepts often compete for five-figure advances — unaffordable for smaller presses. Substantial developmental editing, to frame them accessibly without sacrificing their scholarly underpinnings, is often offered. Publicists work with the news media as well as science journals: publishers hope "for reviews not only in *The Times Literary Supplement*, but also in the broadsheets", says Sophie Goldsworthy, Oxford University Press's editorial director for academic and

"Almost everything we publish gets pirated. In one sense, everything is open-access."

trade books. These books are generally expected to sell anywhere from 1,500 to 5,000 copies, with the exceptional case — such as economist Thomas Piketty's *Capital in the Twenty-First Century* (Belknap/Harvard, 2014) — selling in the millions. E-books have yet to match print books in revenue or unit volume.

Broad changes may shift the traditional publishing structure. University presses and other academic publishing houses have long managed separate businesses for books and journals. Now, in a move towards flexibility in formats, several, including the University of California Press in Oakland, California, are strengthening connections to build an overall publishing strategy — or merging these departments completely. Cambridge delivers its journals and books through a single website, enabling readers of one format to discover products in the other and allowing innovation at a product level.

But such radical change in production and distribution is not easy for academic houses. Elsevier and the other large commercial scientific-monograph publishers succeed because size offers considerable cost savings

as they streamline production and distribution. Such publishers have also established a business model increasingly decoupled from sales for an individual book. Starting with Springer (now Springer Nature, owners of this journal) a decade ago, they moved to a bundled-subscription model for academic libraries. Readers like this model because large quantities of books are available in usable formats such as PDFs, rather than restricted digital-rights-management formats. Publishers like it too, because it eases sales to libraries.

TRANSITION AND TRADE-OFF

All academic publishers have begun to make their books available for digital sale, both through Amazon's Kindle platform and to academic libraries. Smaller presses are more likely to sell to libraries through third-party aggregators such as EBSCO, Project MUSE, ProQuest and JSTOR (I work for Ithaka S+R, whose parent firm operates JSTOR). Several use platforms based on technology from HighWire, Silverchair and Atypon (acquired this year by publisher Wiley of Hoboken, New Jersey).

Others have taken steps to develop or operate their own platforms, supplementing or supplanting these other approaches. It is a way of being "a little more self-directed", as Brand puts it, and enables greater product differentiation and format innovation.

Some efforts to develop new formats are under way. Cambridge has introduced concise, updatable treatments called Elements. The University of California Press has debuted Luminos, an open-access publication programme that allows options for multimedia, live links and annotation to be added to its digital versions. And platforms such as Project MUSE are planning to reach back to authors to experiment with even greater format flexibility.

Common themes emerge when academic publishers talk of pressures and opportunities. As Brand notes, digitized content is ever-harder to control. “Almost everything we publish gets pirated one way or another. In one sense, everything is open-access,” she says. Some publishers are fighting this with lawsuits and authentication technology, but others embrace open access. In addition to open-access monograph models available through Luminos, MUSE and JSTOR, a handful of open-access presses have launched in the past five years, including Amherst College Press in Massachusetts and UCL Press in London. Third-party efforts such as Knowledge Unlatched in Berlin syndicate open access across a group of subscribing libraries.

Of greater concern are ‘evidence-based acquisition’ models. In these, a researcher identifies a desired book; only at the point of initial access is some fee charged to the library. Sales cannot be taken for granted and “every book really needs to sing for its supper now”, says Goldsworthy, emphasizing that Oxford University Press does not make publishing decisions on the basis of anticipated return for individual titles. Scholarly publishers take comfort that the high quality of their content will provide some immunity against sales declines, but many still express concerns that these models will have an impact over time.

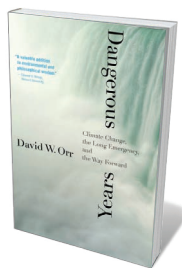
Recent decades have seen a re-sorting of publishers for natural- and social-science books. Some have become expert in building a new business around book publishing, having transitioned from print to digital. They are pursuing subscription-like bundles and open-access models as alternatives to declining library resources for book purchases, and are investing in digital platforms and developing new formats for scholarly communication.

Others are returning to fundamentals. Says Dougherty: “We’re good at taking interesting and important books by scholars and framing and packaging them well for the broader market that we are trying to reach.” With this approach, some university presses have meaningfully advanced public understanding of scholarship while also achieving business success.

But their path ahead is unclear: should reading practices for these materials transition entirely to electronic formats, there will be a substantial reconfiguration as print sales dry up. Over time, issues to do with business models, distribution and scale are likely to emerge as real dilemmas for this small number of presses, and the very important translational books that they publish. ■

Roger Schonfeld is director of Ithaka S+R’s library and scholarly communication programme in New York City. He tweets @rschon.
e-mail: roger.schonfeld@gmail.com

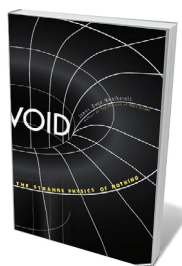
Books in brief



Dangerous Years

David W. Orr YALE UNIVERSITY PRESS (2016)

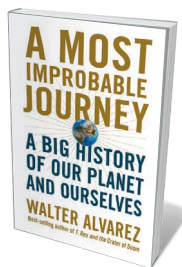
How will we survive the damaging impacts of our high-throughput, extractive economy? Environmental thinker David Orr delivers a finely pointed analysis, ranging over sociopolitical norms and the ethical consciousness needed to transform them. Transitioning to a “sustainable democracy” will demand that we up the ante in areas such as community engagement, science and governance. The shift must be holistic because we are losing the race with technological advance — and, as he notes, “Everything has changed but our manner of thinking, which remains tribal, insular and myopic.”



Void: The Strange Physics of Nothing

James Owen Weatherall YALE UNIVERSITY PRESS (2016)

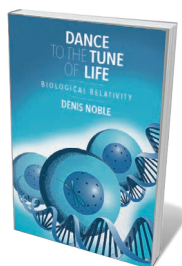
This exploration of empty space — “much ado about nothing”, as philosopher of science James Owen Weatherall has it — is a model of concision. On his tour of physics’s ‘greatest hits’, Weatherall begins with Isaac Newton’s vision of space as nothingness extending in all directions. He then moves through the “aetherial substance” of James Clerk Maxwell, Albert Einstein’s general theory of relativity (which posits a rich structure in space-time), quantum theory (which suggests that the vacuum is a state of matter) and string theory (which is riddled with vacuums). An ode to the plenty of nothing.



A Most Improbable Journey: A Big History of Our Planet and Ourselves

Walter Alvarez W. W. NORTON (2016)

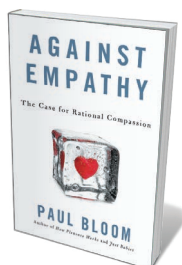
It’s likely that a ‘big history’ tracing the evolution of the Universe and all life sits on a bookshelf near you. Geologist Walter Alvarez joins in with this engaging, yet improbably brief, slalom through the science. Alvarez, who co-originated the impact theory of dinosaur extinction, hinges his primer on contingencies — unpredictable formative cosmic or biological events — and original perspectives. So his discussion of cosmic expansion features Milton Humason rather than co-discoverer Edwin Hubble, whereas his passage on Earth’s formation emphasizes how the planet “makes resources useful”.



Dance to the Tune of Life: Biological Relativity

Denis Noble CAMBRIDGE UNIVERSITY PRESS (2016)

In this audacious riposte to neo-Darwinism and the ‘selfish gene’, systems biologist Denis Noble casts life as emerging from biological processes that operate at various scales and levels. Genes, cells, tissues, organs and organ systems ‘dance’ to the tune of the organism-orchestra. From this he derives his principle of biological relativity: that organisms are “multi-level, open stochastic systems” in which behaviour at all levels is causatively complex. Noble’s narrative is sweeping, covering everything from cosmology and Einstein’s theories of relativity to symbiogenesis and epigenetics.



Against Empathy: The Case for Rational Compassion

Paul Bloom ECCO (2016)

Psychologist and seasoned writer Paul Bloom presents a contrarian characterization of empathy — defined as the ability to feel what you believe others feel — as a wolf in sheep’s clothing. Bloom argues that although it may be seen by some as a “magic bullet of morality”, empathy is actually biased and short-sighted, and can corrode relationships and spark violence. He argues instead for a rational compassion decoupled from the intense shared emotionality of empathy. A nuanced foray into some fraught grey areas. [Barbara Kiser](#)

Correspondence

'Tree of life' took root 150 years ago

As thousands of scientists and policymakers gather in Mexico this month for the COP13 summit on biodiversity (see www.cbd.int/cop2016), we should take a moment to celebrate the earliest 'tree of life' model of biodiversity.

Charles Darwin published the idea of a tree of life in *On the Origin of Species* in 1859. Seven years later, German zoologist Ernst Haeckel painstakingly drew up a much more comprehensive tree (pictured). This represented Earth's wealth of species in the context of evolution — a concept he dubbed phylogeny (*General Morphology of Organisms*; 1866).

The root of the tree symbolizes a common primordial ancestor from which all other forms emerged. Haeckel developed his tree over almost 1,000 pages, basing it on palaeontological, embryological and systemic data — a precursor to modern biology's phylogenetic trees.

He also coined the term ecology ('oecologie'), describing it as "the whole science of the relations of the organism to the environment including, in the broad sense, all the 'conditions of existence'".

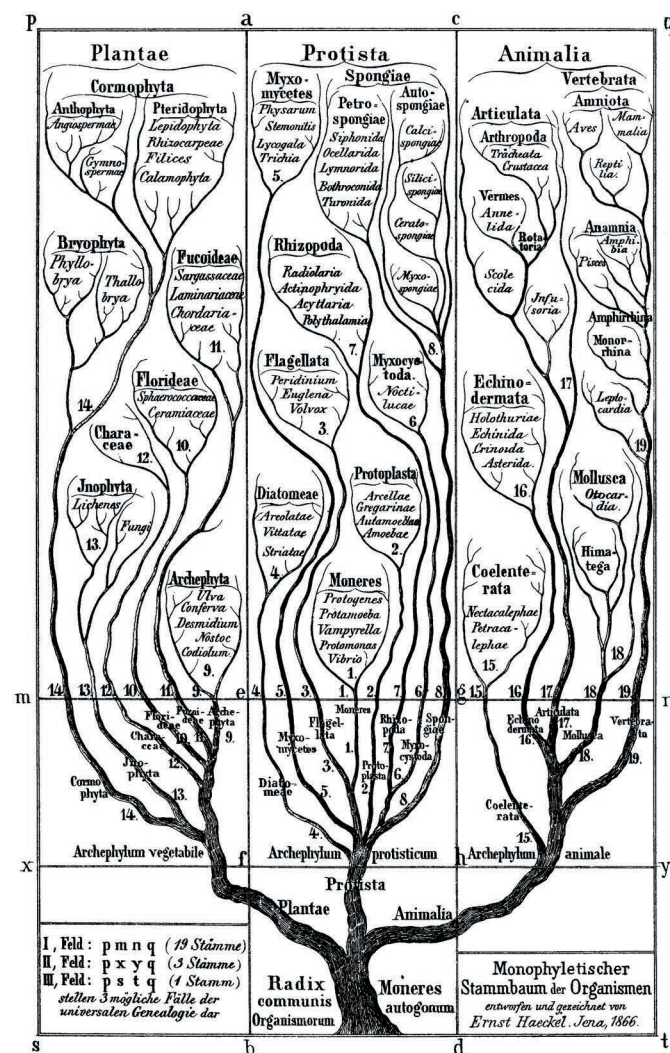
Haeckel's ideas were harbingers for discoveries such as that by ecologist Santiago Soliveres and colleagues, who demonstrated that ecosystem multifunctionality depends on high species richness (*Nature* 536, 456–459; 2016).

Uwe Hossfeld, Georgy S. Levit
Friedrich Schiller University, Jena, Germany; and ITMO University, St Petersburg, Russia.
gslevit@corp.ifmo.ru

Conservation: smart advocacy needs data

We agree that data alone will not save species (A. M. Ellison *Nature* 538, 141; 2016). However, using data in combination with smart advocacy can make a difference.

For example, ecologists



specializing in seabirds, fishes and coral reefs shared their data to show that protection of an area in the northwest Hawaiian Islands was scientifically justified. This evidence was required in order to use the US Antiquities Act to conserve the region. As a result, this became the world's largest no-take protected area earlier this year.

Contrary to Aaron Ellison's implication, some ecologists have been inspirational advocates for biodiversity. For example, biologist E. O. Wilson campaigns to protect half the planet (*Half-Earth*; Norton, 2016) and ecologist Daniel Janzen helped to save huge tracts of dry forest in Costa Rica (W. Allen *Green Phoenix*; Oxford Univ. Press, 2003).

More scientists need to be engaged with society and make the case for conservation — with good, compelling data in hand. None of us wants our publications to serve as pages in a wildlife obituary.

Douglas J. McCauley, Francis H. Joyce University of California, Santa Barbara, USA.
Jane Lubchenco Oregon State University, Corvallis, Oregon, USA.
dmccauley@ucsb.edu

Conservation: big data boost in China

I disagree with Aaron Ellison's contention that biodiversity data only rarely drive conservation decisions (*Nature* 538, 141; 2016).

In China, better data are guiding changes in conservation policies and on-the-ground actions.

For instance, huge long-term data sets on species, ecosystems and human activities have enabled China to identify many priority conservation areas for protecting biodiversity and maintaining ecological security (R. Wu *et al.* *PLoS ONE* 9, e103783; 2014). Intensive development is prohibited in these areas.

Furthermore, a huge amount of data from hundreds of scientists are used by China's Ministry of Environmental Protection to compile biodiversity 'red lists' (R. Wu *Science* 353, 657; 2016). And the systematic collection of more and better data is crucial to implementation of the government's nationwide 'Red Lines', which demarcate ecological conservation regions (see W. Sang and J. C. Axmacher *Nature* 531, 305; 2016).

Ruidong Wu Yunnan University, Kunming, Yunnan, China.
rdwu@ynu.edu.cn

Conservation: thrive on slings and arrows

Aaron Ellison rightly calls for conservation scientists to engage more actively in the political process (*Nature* 538, 141; 2016). Unfortunately, those doing so can become targets for physical and verbal threats, personal abuse, bullying and trolling on social media.

As a conservation scientist, I have experienced many such reactions at first hand. I have been embroiled for more than 30 years in the politics of native forest logging, tree-plantation expansion and culling of over-abundant native and exotic animals. Resilience against such attacks is underpinned by a drive to conduct and communicate high-quality science that contributes to the conservation and management of natural environments and biodiversity.

Communicating good data

and sound science to resource managers, policymakers and politicians is crucial for managing environmental resources and for reinforcing political decisions. Scientists need more training and mentoring in such skills — and in dealing with unwelcome fallout from lobbyists and the public.

David Lindenmayer *Australian National University, Canberra, Australia.*
david.lindenmayer@anu.edu.au

Cities: factor in their biological impact

Richard Forman and Jianguo Wu identify several zones around the world that might be suitable for future urbanization (*Nature* **537**, 608–611; 2016). We suggest that the problem of supporting a growing population needs to be considered from a biological as well as an urban-planning perspective: 'suitable' is not necessarily synonymous with 'biologically sustainable'.

The authors propose that urbanization should avoid biodiversity hotspots — regions that are rich in endemic species and undergoing rapid habitat loss. But this would exclude highly biodiverse yet well-conserved areas, which also need environmental protection.

Forman and Wu also call for global-scale planning. Until that is properly coordinated, governments should not view natural ecosystems as offering potential accommodation for the next billion people. Instead, they need to maximize land-use potential while minimizing its biological impact — for example, by improving cities and the surrounding lands that feed them, and by promoting biodiversity research and protection in conserved areas.

We probably do not need more "suitable" land: we need to make sustainable and efficient use of the land we already live on.

Alexandre Antonelli *University of Gothenburg and Gothenburg Botanical Garden, Sweden.*

Allison Perrigo *Forest Cat Editing, Uppsala, Sweden.*
alexandre.antonelli@bioenv.gu.se

Cities: new fringes to act as safety nets

Richard Forman and Jianguo Wu suggest that sustainable city expansion should be restricted to selected urban peripheries (*Nature* **537**, 608–611; 2016). Our investigations across five Indian cities indicate that proper planning of such peri-urban areas is crucial.

We define peri-urban areas as a 5-kilometre fringe of development beyond a city's jurisdiction. The India Research Initiative for Peri-Urban Human–Animal–Environment Interface (see <http://perimilk.phfi.org>), funded by Canada's International Development Research Centre, is examining the effects of expansion around Bangalore, Ludhiana, Guwahati, Bhubaneswar and Udaipur on city dwellers. The populations and calculated decadal growth rates of these cities vary from 8.44 million and 49% (Bangalore) to 0.45 million and 14% (Udaipur).

Our findings indicate that poor farming practices across all of these peri-urban areas could put food safety at risk, promote the spread of tuberculosis and other zoonotic infections, and accelerate microbial resistance through antibiotic overuse in livestock (our unpublished results). Public health is further threatened by minimal access to proper health care and by inadequate waste-management systems.

Without strict oversight, sound planning and investment in building infrastructure, peri-urban belts stand to aggravate many of the adverse effects of rapid urban growth.

Pranab Chatterjee, Manish Kakkar *Public Health Foundation of India, Gurgaon, India.*

Tamoghna Biswas *Kolkata Medical College, Kolkata, India.*
manish.kakkar@phfi.org

Trump: voters show importance of SDGs

Many of the factors that contributed to Donald Trump's win in the 2016 US presidential election are those that make achievement of the

United Nations' Sustainable Development Goals (SDGs) so pressing. Drawn up to end poverty (goal 1) and inequality (goal 10) and to ensure prosperity, among other aims, the tenet of the SDGs is to 'leave no one behind'.

In the United States, rising inequality and increasing poverty have created a groundswell of the disenfranchised and the dispossessed. Health care (goal 3) and tertiary education (goal 4) are unaffordable for many. Decent work (goal 8) and industry (goal 9) are under threat from automation and globalization, leading to job losses, fuelling racism and driving up crime (counter to goal 11). Extreme weather events related to climate change are on the rise.

All of this has created a potent cocktail of discontent that has manifested in anger towards the establishment. Yet I fear for most, if not all, of the SDGs under a Trump presidency. For example, he has vowed to undermine action against climate change (goal 13) and could jeopardize the development of clean energy (goal 7). He could put biodiversity at risk by making cuts to the Environmental Protection Agency (goals 14 and 15), and he seems to dismiss gender equality (goal 5).

David Griggs *Monash University, Clayton, Australia.*
dave.griggs@monash.edu

Trump: offset fracking risks

Unconventional gas extraction ('fracking') is likely to be central to the energy policy of president-elect Donald Trump. I urge the US Congress to tighten current regulations and strike a balance between energy independence and the safeguarding of public health and the environment (see also J. C. S. Long *Nature* **539**, 495; 2016).

The shale-gas revolution has given the United States a security blanket for energy, at a cost to the environment and to health (see M. L. Finkel and J. Hays *J. Epidemiol. Community Health* **70**, 221–222; 2016). Several European countries and US states (including New York, Maryland and parts of California, Colorado

and Texas) have therefore issued a moratorium on fracking.

This means that the US energy-policy debate needs a strong public-health and environmental presence. Before drilling starts, it is crucial to collect baseline data to track any related increase in morbidity and mortality. During the extraction process, environmental monitoring will be needed to assess air pollution and its impact; and details of all chemicals used in the process must be disclosed. Flowback fluids should be properly disposed of to protect surface and groundwater. These safeguards will help to counter potential harm to human health and the environment.

Madelon Finkel *Weill Cornell Medical College, New York, USA.*
mfinkel404@aol.com

Trump: polls right, models wrong

Contrary to popular opinion, the polls were not wrong in last month's US presidential election (see also *Nature* **539**, 339; 2016). The most recent polls in each state predicted the outcome for individual states — but only when the 95% confidence interval lay outside the polling error. For those predictions with confidence intervals within the polling error, the results were uncertain.

Many pollsters used predictive models to estimate probabilities. Permutations and models are reliable, however, only when projections capture the uncertainty of both the underlying data and the model itself. We cannot claim to make highly certain predictions from highly uncertain data. In this case, some models claimed a certainty of 99% even though 31% of the electoral-college vote was too close to call.

We need to improve the way we explain uncertainty: uncertain data are not wrong, only uncertain. And we must become better consumers of models.

The polls were right. The models were wrong. In this era of big data, we need to emphasize the distinction.

Brandon Steelman *Clear Gene, Moss Beach, California, USA.*
steelman@cleargenes.com

Susan Lee Lindquist

(1949–2016)

Biologist who found unexpected power in protein folding.

Susan Lee Lindquist revealed the profound ramifications of a process that most scientists had overlooked: protein folding. Through elegant experiments in yeast, plants, flies and human cells, Lindquist demonstrated how this process by which proteins adopt their proper shapes fuels evolution. It can buffer the effects of genetic variation, allow new traits to emerge and enable the rapid evolution of new adaptations. Her insights have paved the way for innovative strategies to treat diseases including Alzheimer's, Parkinson's and cancer.

Lindquist was a visionary who connected concepts across disparate disciplines. When her bold ideas were met with doubt, she persevered, gathering evidence until she changed biological thinking. Her positivity and love for discovery were infectious.

Lindquist, who died from cancer on 27 October, was born in Chicago, Illinois, in 1949. Early on, she became enthralled by a teacher's question: what is life? In search of answers, she studied microbiology at the University of Illinois Urbana–Champaign. She earned her PhD in molecular and cellular biology from Harvard University in Cambridge, Massachusetts, in 1976.

It was at Harvard that she began studying how cells respond to heat and other stresses by producing heat-shock proteins (Hsps). After a brief postdoc at the University of Chicago, she joined the faculty there in 1978 and became a full professor in 1988. In 2001, she moved to the Whitehead Institute for Biomedical Research at the Massachusetts Institute of Technology in Cambridge. She served as its director until 2004, and as a professor of biology from 2001 until her death.

Lindquist focused on Hsp90. This abundant protein acts as a molecular chaperone, helping to fold many proteins important for cell signalling. She identified two surprising roles for Hsp90 in evolution. First, she found that Hsp90 could protect cell-signalling pathways from the effects of mutations in other proteins. If Hsp90 function was compromised by environmental stress, genetic variants would fold differently and new traits would appear rapidly. Her striking experiments in plants, for example, revealed that inhibiting Hsp90 function could yield some beneficial phenotypes, including altered growth of stems and roots, and increased resistance to herbivores.

Second, Lindquist established that Hsp90



can allow mutations that would otherwise result in an unstable protein to produce a functioning protein with a new phenotype. Both roles expand the repertoire of traits that can be selected through evolution. Lindquist demonstrated that these powerful mechanisms are probably universal across eukaryotes.

Lindquist also shattered dogma about misfolded proteins associated with disease. She discovered that Hsp104 could wrest clumps of misfolded proteins apart and return them to their functional form. This finding led her to study prions. These infectious proteins cause normally folded proteins to assume self-replicating conformations that cause fatal neurodegenerative disorders such as Creutzfeldt–Jakob disease and scrapie.

She was fond of saying that prions transmit perduring molecular memories. At a time when prions were deemed merely disease-causing villains, she established that they can also confer beneficial functions, including antibiotic resistance and improved metabolism. Indeed, Lindquist helped to reveal how prions may function in long-term memories by stimulating translation at synapses.

Beginning in the 1990s, Lindquist championed yeast as a model to probe both cancer and neurodegeneration. Sceptics questioned (and still do) why anyone would try to study a brain disease in yeast. Undaunted, she developed a series of yeast models that express each of the major

human neurodegenerative disease proteins. As in the brain, these proteins formed aggregates and were toxic in yeast. She identified yeast genes that reduced the toxicity of protein misfolding connected to Parkinson's disease and established that these same genes are protective in mice and rats.

She co-founded two Cambridge companies, FoldRx Pharmaceuticals (acquired by Pfizer in 2010) and Yumanity Therapeutics to uncover therapies for neurodegenerative diseases caused by protein misfolding. Her use of yeast in drug screening is now accepted practice.

Researchers from diverse backgrounds — from physicists to physicians — wanted to work in Lindquist's lab. Both of us pursued postdocs at the Whitehead after being inspired by seminars she gave at our graduate institutions. One of us (J.S.) moved from the United Kingdom to the United States and the other (A.D.G.) shifted discipline to work with her.

Sue believed that scientists had a moral obligation to address important societal problems. In her papers, she used language that was accessible to wide audiences. She was a nurturing mentor, inviting group members to her house on weekends to write manuscripts. We were fortunate to get to know her in this personal setting, interacting with her wonderful daughters and husband, and learning that it was possible to be a world-class scientist and have a family. Her unwavering commitment to promoting women in science is memorialized by the Whitehead Institute Fund to Encourage Women in Science.

Somehow, Sue found time in her unrelenting schedule to sit on the grass and watch her students and postdocs play soccer on Friday afternoons in Cambridge — always there to cheer us on. The world feels smaller without her. ■

James Shorter is associate professor of biochemistry and biophysics at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. **Aaron D. Gitler** is associate professor of genetics at Stanford University, Stanford, California, USA. Both worked with Susan Lindquist in the 2000s as postdocs at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts.
e-mails: jshorter@mail.med.upenn.edu, agitler@stanford.edu

CEAL CAPISTRANO

Consistency in large pharmacogenomic studies

ARISING FROM B. Haibe-Kains *et al.* *Nature* **504**, 389–393 (2013); doi:10.1038/nature12831

Haibe-Kains *et al.*¹ reported inconsistency between two large-scale pharmacogenomic studies—the Cancer Cell Line Encyclopedia (CCLE)² and the Cancer Genome Project (CGP)³. Upon careful analysis of the same data, we have come to quite different and much more positive conclusions. Here we highlight the most important reasons for this. There is a Reply to this Comment by Safikhani, Z. *et al.* *Nature* **540**, <http://dx.doi.org/10.1038/nature19839> (2016).

To assess the concordance of two large studies of the efficacy of cancer drugs, Haibe-Kains *et al.*¹ compared the correlation in drug sensitivity measures with correlation in gene expression measured on the same human cancer cell lines. The authors reported correlation ‘between’ cell lines for gene expression but, inconsistently, ‘across’ cell lines for drug sensitivity (see Methods). On re-analysis, we found much higher correlations between cell lines than across cell lines for both gene expression and drug sensitivity measures (median Spearman’s rank correlation coefficient (r_s) = 0.88 between cell lines, r_s = 0.56 across cell lines for expression; median r_s = 0.62 between cell lines and r_s = 0.35 across cell lines for area under the curve (AUC), a drug sensitivity measure). Thus, by correcting this inconsistency, the correlations for expression and drug sensitivity data were far more similar than was originally reported, which markedly undermines the authors’ interpretation of the relative quality of expression and drug sensitivity datasets.

In addition, the fundamental issue is that the authors’ reported Spearman’s correlation coefficients do not fairly reflect the concordance of drug sensitivity between the studies, because of the lack of variability in drug response, which arises owing to the highly targeted nature of many of the drugs assessed. To see why correlation is not an appropriate measure of biological concordance for these data, consider the hypothetical example of a drug that is not effective against any cell lines, which is a possibility for an experimental drug. In such a case, the randomly fluctuating measurement error, which is inherent in biological assays, will dominate over the non-existent biological variability, meaning that there could be no expectation of correlation between repeated measures of drug sensitivity—assuming other experimental variables are held constant. In this study, many of the drugs were highly targeted agents, which by design require specific, and often rare, molecular targets for response (see Supplementary Table 1). Consider nilotinib, which targets the *BCR-ABL1* fusion gene and was suggested in ref. 1 to exhibit ‘poor consistency’ between CGP and CCLE (r_s = 0.1 for AUC). In CGP, *BCR-ABL1* status was reported to be strongly associated with drug sensitivity (P = 2.54×10^{-65}), accurately reflecting the known biology. *BCR-ABL1* status was not reported by CCLE; however, upon re-analysis we identified three *BCR-ABL1*-positive cell lines among the 189 nilotinib-treated cell lines that overlapped CGP, and these were also the three most sensitive samples (P = 9×10^{-7}). Hence, despite the fact that these drug sensitivity data were accurately recapitulating biological expectations in both studies, the authors’ criteria classified nilotinib sensitivity incorrectly. Of the 577 cell lines screened in CGP, 573 do not contain the nilotinib target, that is, the *BCR-ABL1* fusion gene. Thus, given (as expected) no drug response in almost all cell lines screened (median AUC across all cell lines = 0.99; AUC of 1 represents no drug response; Fig. 1a, Supplementary Table 1), there was little biological variability across most of the cell lines, resulting in low correlation between the repeated measurements made by CCLE and

CGP, despite clearly concordant results. Similarly, most other drugs that the authors compared were also targeted agents, meaning that this lack of drug response was common; for 10 of the 15 drugs, the median

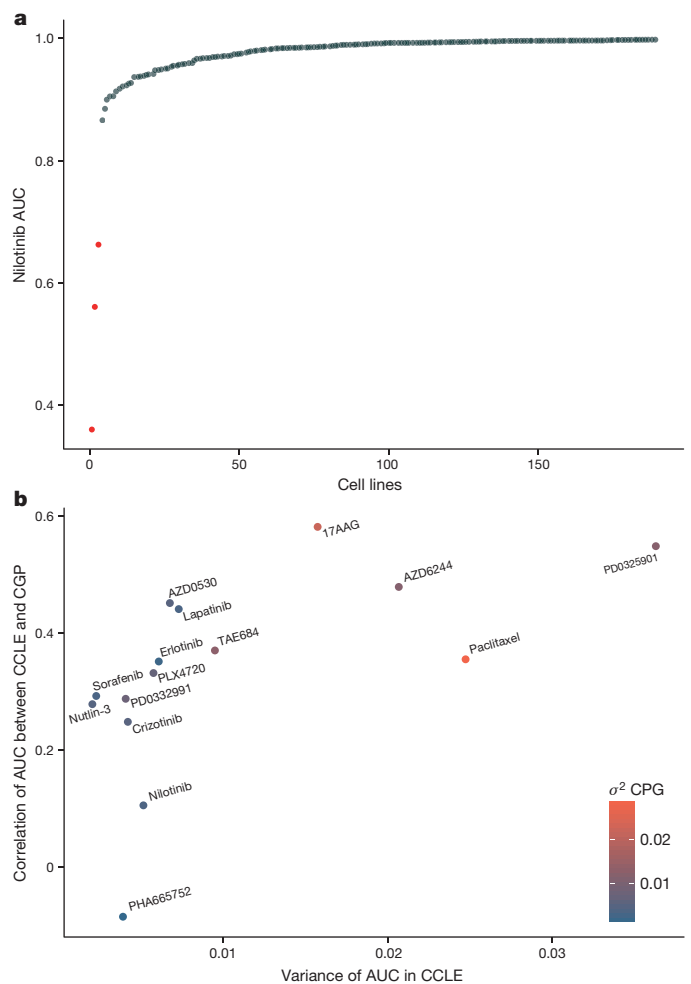


Figure 1 | Limitations of using a correlation metric for the assessment of concordance between the CGP and CCLE drug sensitivity data.

a, Highly targeted agents (such as nilotinib) highlight a major limitation of the test for concordance reported in ref. 1. Scatterplot shows the nilotinib AUC values (in CGP) for the 189 cell lines that were screened by both CGP and CCLE. Only a very small proportion of cell lines achieve a response, that is the three *BCR-ABL1*-positive cell lines highlighted in red. This almost complete lack of biological variability renders a Spearman correlation ineffective as a means to assess concordance. **b**, The test for concordance in ref. 1 is confounded by variability in the drug response. Scatterplot shows the strong association between ‘Spearman’s correlation of AUC between CCLE and CGP’ and ‘variance of AUC in CCLE’. Drugs with a more variable AUC value are more likely to be highly correlated between CCLE and CGP (r_s = 0.83, P = 1.9×10^{-4}). The points have been colour coded by their variance of AUC in CGP, which is also significantly associated with both ‘variance of AUC in CCLE’ and ‘Spearman’s correlation of AUC between CCLE and CGP’. σ^2 CPG denotes the variance of AUC in CGP.

AUC was greater than 0.90 in CGP, and 8 of these 10 also have median AUC values greater than 0.9 in CCLE, resulting in little variability across most cell lines when treated with these drugs. We identified a systematic relationship between variability in drug response in either study and correlation between the two studies (Fig. 1b). A valid comparison of CGP and CCLE should consider the pharmacology of the drugs screened and in particular the differences in the variability induced by different drugs. Nilotinib was not an isolated case; despite the highly experimental nature of many of the compounds screened by CCLE and CGP, we still identified several expected associations that were consistently reported by both studies, including *ERBB2* for lapatinib⁴, *NQO1* expression for 17-AAG⁵, *BRAF* mutation for PD-0325901 (ref. 6), AZD6244 (ref. 7), and PLX4720 (ref. 8), *MDM2* for nutlin-3a (ref. 9), and *MET* for crizotinib¹⁰ (Supplementary Table 1). Finally, the utility of these pharmacogenomic datasets is now further supported by the findings that models fit using data from CGP could reliably predict drug response in several clinical trials^{11,12}.

In summary, our analysis shows that the conclusions of Haibe-Kains *et al.*¹ are unsubstantiated, and we propose that a fair assessment of concordance between large pharmacogenomic datasets will require the development or adaptation of methods that account for the issues raised here, although great care will be required to ensure that such methods do not introduce their own unforeseen biases.

Methods

In CGP and CCLE, using ordered data common to both studies, gene expression and drug sensitivity (AUC) values can be arranged in $n_1 \times m$ and $n_2 \times m$ matrices, respectively, in which m is the number of cell lines, n_1 is the number of genes and n_2 is the number of drugs common to both studies. Correlations 'between' cell lines are calculated by the correlation of matching columns of CGP and CCLE matrices (vectors of length n_1 for expression or n_2 for AUC). Correlations 'across' cell lines are the correlations of matching rows (vectors of length m for both data).

To achieve easy reproduction of our results, we have made the source code for our analysis available in a GitHub repository (https://github.com/paulgeeheher/nature_bca).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Paul Geeleher¹, Eric R. Gamazon^{1,2,3}, Cathal Seoighe⁴,
Nancy J. Cox^{1,2} & R. Stephanie Huang¹

Safikhani *et al.* reply

REPLYING TO P. Geeleher, E. R. Gamazon, C. Seoighe, N. J. Cox & R. S. Huang *Nature* **540**, <http://dx.doi.org/10.1038/nature19838> (2016)

In the accompanying Comment¹, Geeleher *et al.* claim to have discovered overall consistency between the Cancer Genome Project (CGP)² and the Cancer Cell Line Encyclopedia (CCLE)³ by analysing the response of three cell lines containing the *BCR-ABL1* fusion gene to the highly targeted drug nilotinib (Supplementary Fig. 1). They use this example to argue that owing to the targeted nature of many of the 15 drugs screened in both CGP and CCLE (Supplementary Table 1), the rest of the drugs should show consistent sensitivity measurements if analysis is limited to the highly sensitive cell lines. However, as we describe in more detail below, the consistency seen in the sensitivities of cell lines to nilotinib is an isolated example that does not generalize to other targeted drugs, supporting our initial finding⁴ of a broader inconsistency in reported phenotypes between CGP and CCLE⁴.

¹Department of Medicine, The University of Chicago, Chicago, Illinois 60637, USA.

email: nancy.j.cox@vanderbilt.edu

email: rhuang@medicine.bsd.uchicago.edu

²Division of Genetic Medicine, Vanderbilt University, Nashville, Tennessee 37232, USA.

³Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands.

⁴Department of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland.

Received 28 March 2014; accepted 22 July 2016.

1. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).
2. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
3. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
4. Konecny, G. E. *et al.* Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res.* **66**, 1630–1639 (2006).
5. Kelland, L. R., Sharp, S. Y., Rogers, P. M., Myers, T. G. & Workman, P. DT-Diaphorase expression and tumor cell sensitivity to 17-allylamino, 17-demethoxygeldanamycin, an inhibitor of heat shock protein 90. *J. Natl. Cancer Inst.* **91**, 1940–1949 (1999).
6. Solit, D. B. *et al.* BRAF mutation predicts sensitivity to MEK inhibition. *Nature* **439**, 358–362 (2006).
7. Dry, J. R. *et al.* Transcriptional pathway signatures predict MEK addiction and response to selumetinib (AZD6244). *Cancer Res.* **70**, 2264–2273 (2010).
8. Tsai, J. *et al.* Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antitumor activity. *Proc. Natl. Acad. Sci. USA* **105**, 3041–3046 (2008).
9. Müller, C. R. *et al.* Potential for treatment of liposarcomas with the MDM2 antagonist Nutlin-3A. *Int. J. Cancer* **121**, 199–205 (2007).
10. Timm, A. & Kolesar, J. M. Crizotinib for the treatment of non-small-cell lung cancer. *Am. J. Health Syst. Pharm.* **70**, 943–947 (2013).
11. Geeleher, P., Cox, N. J. & Huang, R. S. Clinical drug response can be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines. *Genome Biol.* **15**, R47 (2014).
12. Falgreen, S. *et al.* Predicting response to multidrug regimens in cancer patients using cell line experiments and regularised regression models. *BMC Cancer* **15**, 235 (2015).

Supplementary Information is available in the online version of the paper.

Author Contributions P.G. and E.R.G. performed analysis. P.G. drafted the paper. All authors discussed and decided on the scope and technical details of the project, edited and approved the manuscript. R.S.H. and N.J.C. supervised the project.

Competing Financial Interests Declared none.

doi:10.1038/nature19838

concentration (IC₅₀) and area under the curve (AUC)) values in all comparisons (Wilcoxon rank sum test $P < 0.002$). Consequently, our original conclusion that gene expression data are significantly more correlated than pharmacological response data still holds. In fact, the lower correlation of gene expression values across cell lines could suggest that there is even less consistency between the CCLE and CGP studies than we initially reported.

Second, the argument of Geeleher *et al.*¹ that the lack of variability in drug sensitivity measurements may complicate biologically meaningful assessment of concordance between pharmacogenomic datasets is not new, as it was already discussed by our group⁵ and others^{6,7}. Geeleher *et al.*¹ focus on the sole example of nilotinib, for which there are three highly sensitive cell lines out of the 200 cell lines screened in both datasets. However, even among these three cell lines, the AUC values are not concordant; the least sensitive of the three cell lines in CGP is actually the most sensitive one in CCLE (Supplementary Fig. 1 and Supplementary Information). Therefore, the only way to consider these results to be concordant is to classify these three cell lines as sensitive and the remainder as resistant, which cannot easily be done using the waterfall method described in the CCLE study^{3,4}.

The authors claim that nilotinib is not an isolated example¹, but do not seem to propose any statistic to quantify the consistency of the other drugs. We adapted the Matthews correlation coefficient⁸ (AMCC; see Supplementary Methods) to select the optimal cutoff for consistency between drug sensitivity calls, in which only a few cell lines may be sensitive, or between gene expression data, in which the gene of interest may be rarely expressed. As expected, nilotinib yielded an AMCC value of 1, which denotes perfect consistency between the two studies (Supplementary Fig. 3). However, the other drugs yielded much lower AMCC values, with only AZD0530, lapatinib and crizotinib yielding AMCC values of around 0.65, with another five drugs (17-AAG, AZD6244, erlotinib, PD-0325901 and PLX4720) yielding moderate consistency ($0.5 \leq \text{AMCC} < 0.6$), and the rest of the drugs yielding poor consistency ($\text{AMCC} < 0.5$). Notably, there was no systematic relationship between variability in drug sensitivity and AMCC estimates (Supplementary Figs 4 and 5), suggesting that AMCC values, although potentially overoptimistic, are a more appropriate statistic for consistency than Spearman's or Pearson's correlation coefficients. It should be noted that the inter-laboratory replicates of the measurements of camptothecin and AZD6482 sensitivity performed using the same experimental protocols at two different locations within CGP yielded AMCC values of only 0.55 and 0.41 (Supplementary Fig. 6), indicating a lack of reproducibility of drug phenotype measures between biological replicates. Consistent with our previous report, gene expression data yielded significantly larger AMCC than drug sensitivity data across cell lines (Wilcoxon rank sum test $P < 0.006$; Supplementary Fig. 7). This re-analysis confirms that nilotinib is an anecdotal case, and that drug sensitivity measurements for the rest of the drugs (cytotoxic or targeted) remain only poorly to moderately consistent.

Geeleher *et al.*¹ also state that their previous findings⁹ support the utility of the data from the CCLE and CGP, and suggest that their findings provide evidence for a consistency between CGP and CCLE. However, a true test of this assertion would be to train their models on the CGP and to use these to predict phenotypes reported by CCLE (and vice versa). If they could predict the drug response phenotype in the independent validation set with high accuracy, this would provide some quantitative evidence of a consistency between the two datasets in the context of their predictive models¹. However, we¹⁰ and others^{11–13} have shown that such an analysis does not yield robust predictions for most drugs.

In our original report⁴, we found statistically significant non-zero correlations between phenotype measurements for almost all drugs, supporting the fact that there is biologically relevant signal in these

datasets, albeit confounded by significant noise. Concurring with recent reports^{6,7,14}, we identified known gene–drug associations that are reproducible between CGP and CCLE (Supplementary Information); however, half of the known associations were inconsistent across datasets (significant in only one dataset; Supplementary Table 2). In a recent re-analysis of the updated CGP and CCLE datasets, we reported that the discovery of new, potentially weaker biomarkers, which was the main goal of the CGP and CCLE studies, was much more challenging owing to inconsistency in pharmacological phenotypes⁵. In our original conclusions⁴, we argued that additional work is necessary to improve the consistency of phenotypic measures with the ultimate goal of making data from these large-scale projects more useful for development of robust predictors of drug response, and we believe that this conclusion still holds upon our re-analysis. We and others are actively working on identifying stable measures that could lead to improved consistency across inter-laboratory experiments.

Like Geeleher *et al.*¹, we originally hoped to use the CCLE and CGP to develop robust biomarkers that could predict responses to treatment. Although we could use methods to find some consistency in selected subsets of the data, we found no general methods that identified an overall consistency between the studies. Geeleher *et al.*¹ have shown that if the drug sensitivity data are appropriately discretized, consistency can be found for nilotinib—but not for the other compounds tested in both studies. Despite the single example of nilotinib, we conclude as we did originally, that sensitivity phenotypes lack consistency for most of the drugs screened both in CGP and CCLE.

The methods are described in details in the Supplementary Information. The PharmacGx package is open-source and publicly available on Bioconductor (<https://bioconductor.org/packages/PharmacGx/>). The code and associated files required to reproduce this analysis are publicly available on the *cdmug-rebuttals* GitHub repository (<https://github.com/bhklab/cdmug-rebuttals>). The procedure to setup the software environment and run our analysis pipeline is provided in Supplementary Information. This work complies with the guidelines proposed previously¹⁵ in terms of code availability and replicability of results.

Author A. C. Jin was a student in A.H.B.'s laboratory and left shortly after publication of ref. 4, and did not participate in the writing of this Reply. Authors Z.S., P.S. and M.F. developed the PharmacGx software package, which enabled the analyses presented here; A.G. helped with the development of the AMCC statistics, and participated in the interpretation of the results and writing of this Reply.

Zhaleh Safikhani^{1,2}, Nehme El-Hachem³, Petr Smirnov¹, Mark Freeman¹, Anna Goldenberg^{4,5}, Nicolai J. Birkbak^{6,7}, Andrew H. Beck^{8,9}, Hugo J. W. L. Aerts^{9,10,11}, John Quackenbush^{10,12} & Benjamin Haibe-Kains^{1,2,5,13}

¹Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario M5G 2M9, Canada.

email: bhaibeka@uhnresearch.ca

²Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada.

³Institut de recherches cliniques de Montréal, Montreal, Quebec H2W 1R7, Canada.

⁴Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada.

⁵Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada.

⁶The Francis Crick Institute, University College London, London NW1 1AT, UK.

⁷University College London Cancer Institute, London, WC1E 6BT, UK.

⁸Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA.

⁹Harvard Medical School, Boston, Massachusetts 02115, USA.

¹⁰Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.

¹¹Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

¹²Harvard School of Public Health, Boston, Massachusetts 02115, USA.

¹³Ontario Institute of Cancer Research, Toronto, Ontario M5G 1L7, Canada.

1. Geeleher, P., Gamazon, E. R., Seoighe, C., Cox, N. J. & Huang, R. S. Consistency in large pharmacogenomic studies. *Nature* **540**, <http://dx.doi.org/10.1038/nature19838> (2016).
2. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
3. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
4. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).
5. Safikhani, Z. *et al.* Revisiting inconsistency in large pharmacogenomic studies. *F1000Research* <http://dx.doi.org/10.12688/f1000research.9611.1> (2016).
6. Haverty, P. M. *et al.* Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* **533**, 333–337 (2016).
7. The Cancer Cell Line Encyclopedia Consortium and Genomics of Drug Sensitivity in Cancer Investigators. Pharmacogenomic agreement between two cancer cell line data sets. *Nature* **528**, 84–87 (2015).
8. Matthews, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **405**, 442–451 (1975).
9. Geeleher, P., Cox, N. J. & Huang, R. S. Clinical drug response can be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines. *Genome Biol.* **15**, R47 (2014).
10. Papillon-Cavanagh, S. *et al.* Comparison and validation of genomic predictors for anticancer drug sensitivity. *J. Am. Med. Inform. Assoc.* **20**, 597–602 (2013).
11. Dong, Z. *et al.* Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC Cancer* **15**, 489 (2015).
12. Jang, I. S., Neto, E. C., Guinney, J., Friend, S. H. & Margolin, A. A. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. *Pac. Symp. Biocomput.* **2014**, 63–74 (2014).
13. Cortés-Ciriano, I. *et al.* Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. *Bioinformatics* **32**, 85–95 (2015).
14. Goodspeed, A., Heiser, L. M., Gray, J. W. & Costello, J. C. Tumor-derived cell lines as molecular models of cancer pharmacogenomics. *Mol. Cancer Res.* **14**, 3–13 (2015).
15. Sandve, G. K., Nekrutenko, A., Taylor, J. & Hovig, E. Ten simple rules for reproducible computational research. *PLoS Comput. Biol.* **9**, e1003285 (2013).

Supplementary Information is available in the online version of the paper.

doi:10.1038/nature19839

Consistency in drug response profiling

ARISING FROM B. Haibe-Kains *et al. Nature* **504**, 389–393 (2013); doi:10.1038/nature12831

The comparative analysis by Haibe-Kains *et al.*¹ concluded that data from two large-scale studies of cancer cell lines^{2,3} showed highly discordant results for drug sensitivity measurements, whereas gene expression data were reasonably concordant. Here, we cross-compared the two original datasets^{2,3} against our own data of drug response profiles in overlapping cancer cell line panels. Our results indicate that it is possible to achieve concordance between different laboratories for drug response measurements by paying attention to the harmonization of assays and experimental procedures. There is a Reply to this Comment by Safikhani, Z. *et al. Nature* **540**, <http://dx.doi.org/10.1038/nature20172> (2016).

Haibe-Kains *et al.*¹ reported on a comparative evaluation of two drug sensitivity and molecular profiling datasets, one from the Cancer Genome Project (CGP)² and the other from the Cancer Cell Line Encyclopedia (CCLE)³. In their analyses, gene expression profiles between hundreds of common cancer cell lines across all genes showed high consistency between the two studies (median rank correlation (MRC) = 0.85), whereas the drug response data for 15 common compounds were highly discordant (MRC = 0.28 for half-maximum inhibitory concentration (IC₅₀) values). This report¹ and the accompanying commentary⁴ suggested that differences in laboratory protocols, compounds and their tested concentration ranges, and computational methods may account for the differences, but these reports did not elaborate which of these factors are important and whether they can be controlled for.

Here, we reanalysed the dose–response data from both CGP and CCLE using a standardized area under the curve (AUC) response metric, which we call the drug sensitivity score (DSS)⁵. We then compared the CGP and CCLE data with a new dataset of drug responses profiled using the Institute for Molecular Medicine Finland (FIMM) compound testing assay⁶, covering 308 drugs across 106 cancer cell lines. The FIMM data included 45 compounds in common with

CGP and 14 with the CCLE in 50 cell lines (Supplementary Data 1). In the AUC calculation, we unified the drug concentration ranges across the CGP, CCLE and FIMM assays. We observed a significantly higher level of consistency ($P = 4.2 \times 10^{-5}$), especially between the CCLE and FIMM drug response data (MRC = 0.74), as compared to the consistency between FIMM and CGP data (MRC = 0.54) (Fig. 1a).

Similar experimental protocols were applied at FIMM and CCLE, including the same readout (CellTiter-Glo, Promega), similar controls (vehicle as negative control and positive controls of toxic compounds 100 μ M benzethonium chloride or 1 μ M MG132). However, there were also differences, such as the plate format used (1,536 versus 384 wells). Importantly, there was no effort made to standardize cell numbers used or any other parameters between the three laboratories, such as the source, passage number and media used for cells, nor the origin and handling of drugs. Therefore, this observed level of drug response agreement could be substantially improved by further standardization of the laboratory protocols. The CGP experimental protocol differed from the two others in terms of the readout (fluorescent nucleic acid stain Syto 60, Life Technologies), in the use of controls (drug-free cells as negative and no cells as positive controls), and the plate format used (96- or 384-well plates).

We compared the drug response profiles between the same cell lines from different laboratories, in line with the approach of Haibe-Kains *et al.*¹, in which they showed consistency in gene expression profiles from CGP and CCLE (MRC = 0.85)¹. The Haibe-Kains *et al.*¹ approach, in which the correlation is calculated for each drug separately across the cell lines, showed more variability (Fig. 1b), owing to the fact that some drugs show minimal efficacy in all the tested cell lines. Analogously, gene expression correlations vary more widely when analysed at the level of genes across cell lines (MRC = 0.58 between CGP and CCLE), as certain genes are not expressed above technical noise. Although both ways to compare the data are relevant to the overall goal of personalized

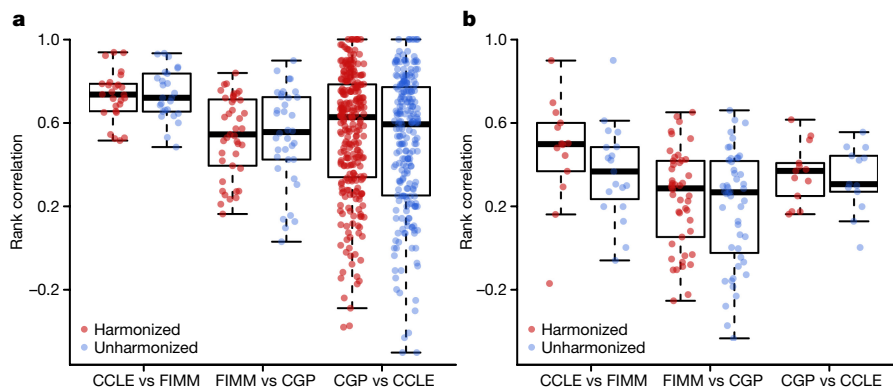


Figure 1 | Consistency between drug response profiles across FIMM, CCLE and CGP. **a**, In the ‘between’ cell line comparison, the Spearman rank correlation coefficient was calculated for the pairwise overlapping cell lines and over shared sets of compounds between FIMM–CCLE (26 cell lines, 14 compounds), FIMM–CGP (41 cell lines, 45 compounds), and CCLE–CGP (268 cell lines, 13 compounds). The drug sensitivity score (DSS) profiles between FIMM and CCLE showed improved correlation compared with the correlation between FIMM versus CGP or between CGP versus CCLE ($P = 4.2 \times 10^{-5}$ or $P = 0.0045$, respectively, two-sided Wilcoxon rank-sum test). **b**, In the ‘across’ cell line comparison¹, the

correlation was calculated for each drug across all of the cell lines. This led to decreased consistency, yet still showed the same inter-laboratory differences after standardized data analysis ($P = 0.00034$, two-sided Wilcoxon rank-sum test, for comparisons both between FIMM versus CCLE and FIMM versus CGP or CGP versus CCLE correlations). The coloured points indicate whether the comparison made use of the harmonized data analysis (see Methods), or the AUC values provided in the original studies^{1,2}. The box and horizontal bar represent the interquartile range and median of the correlation coefficients, respectively, and the whiskers denote the most extreme data points.

therapy, emphasized in the original publications^{2,3}, the same evaluation approach should ideally be used when comparing the consistency of gene expression and drug response measurements.

In summary, we show that standardization of assay methods and laboratory conditions will help to improve the inter-laboratory agreements in drug response profiling. Global standards, similar to the minimum information about a microarray experiment (MIAME) standard for the microarray data⁷, should be developed.

Methods

We scaled the drug response readouts using the available positive and negative controls from CGP, CCLE and FIMM screens (FIMM dose–response data points are provided in Supplementary Data 2). Curve fitting for the scaled dose–response curves was based on the original dose ranges, using the four-parameter logistic model and the Levenberg–Marquardt algorithm^{8,9}. We used the median of the triplicate measurements from CCLE (<http://www.broadinstitute.org/ccle/home>), and the single dose–response measurements from CGP (provided by M. Garnett). In the harmonized data analysis, DSS was calculated based on the fitted dose–response models (R-package available at <https://bitbucket.org/BhagwanYadav/drug-sensitivity-score-dss-calculation>)⁵. The DSS integration was restricted to the common concentration window shared between CGP, CCLE and FIMM, which was limited by the narrower dose range of the CGP data. The compounds were identified and matched using InChIKeys¹⁰. Cell line matching was based on the named mapping file provided by Haibe-Kains *et al.*¹, followed by manual curation to guarantee that we compared exactly the same cell lines and chemical compounds.

This work was financially supported by Academy of Finland (grants 269862 and 279163, T.A.; 277293, K.W.; Center of Excellence for Translational Cancer Biology, O.K.), Cancer Society of Finland (K.W., O.K., T.A.), JSPS Bilateral programs (A.H.), Sigrid Juselius Foundation (A.H., K.W., O.K.), European Union's Seventh Framework Programme (FP7, grant 258068, EU-FP7-Systems Microscopy, O.K.), and Doctoral Programme in Biomedicine (J.P.M. and B.Y.).

John Patrick Mpindi¹, Bhagwan Yadav¹, Päivi Östling^{1,2}, Prson Gautam¹, Disha Malani¹, Astrid Murumägi¹, Akira Hirasawa^{1,3}, Sara Kangaspeska¹, Krister Wennerberg¹, Olli Kallioniemi^{1,2} & Tero Aittokallio^{1,4}

¹Institute for Molecular Medicine Finland (FIMM), University of Helsinki, 00290 Helsinki, Finland.

email: tero.aittokallio@fimm.fi

²Science for Life Laboratory, Department of Oncology and Pathology, Karolinska Institutet, 17165 Solna, Sweden.

³Department of Obstetrics & Gynecology, Keio University School of Medicine, 160-8582, Tokyo, Japan.

⁴Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland.

Received 4 April 2014; accepted 12 October 2016.

1. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).
2. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
3. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
4. Weinstein, J. N. & Lorenzi, P. L. Cancer: Discrepancies in drug sensitivity. *Nature* **504**, 381–383 (2013).
5. Yadav, B. *et al.* Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Sci. Rep.* **4**, 5193 (2014).
6. Pemovska, T. *et al.* Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. *Cancer Discov.* **3**, 1416–1429 (2013).
7. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29**, 365–371 (2001).
8. Levenberg, K. A method for the solution of certain problems in least squares. *Q. Appl. Math.* **2**, 164–168 (1944).
9. Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **11**, 431–441 (1963).
10. Nicola, G., Liu, T. & Gilson, M. K. Public domain databases for medicinal chemistry. *J. Med. Chem.* **55**, 6987–7002 (2012).

Supplementary Information is available in the online version of the paper.

Author Contributions J.P.M. and B.Y. performed computational experiments and analysed data; these authors contributed equally to this work. P.G., D.M., A.M., A.H. and S.K. designed and performed drug response experiments. J.P.M., B.Y., P.Ö., O.K. and T.A. designed the study, interpreted data and wrote the manuscript. K.W. provided critical review and comments on the study and manuscript. O.K. and T.A. supervised the study.

Competing Financial Interests Declared none.

doi:10.1038/nature20171

Safikhani *et al.* reply

REPLYING TO J. P. Mpindi *et al.* *Nature* **540**, <http://dx.doi.org/10.1038/nature20171> (2016)

The accompanying Comment¹ by Mpindi *et al.* is an important contribution to the discussion of pharmacogenomic consistency for several reasons. Mpindi *et al.*¹ were able to reproduce our initial finding² of a substantial inconsistency between the pharmacological profiles generated within the Cancer Genome Project (CGP)³ and the Cancer Cell Line Encyclopedia (CCLE)⁴, and explored potential reasons behind the problem and possible solutions. To do this, they compared the CGP and CCLE to a new dataset generated by the Institute for Molecular Medicine Finland (FIMM) that includes 308 drugs that were tested across 106 cancer cell lines⁵. The authors shared the subset of the FIMM data overlapping with CGP and CCLE, including 52 drugs tested in up to 50 cell lines (drug dose–response curves and their comparison with CGP and CCLE curves are available in Supplementary Data). Overall, their comparative analysis¹ of this newly released dataset supports our published finding² of greater consistency between studies in which there is similarity in experimental methods. We agree with Mpindi *et al.*¹ that harmonizing the readout, drug concentration range, and statistical estimator makes it possible to

achieve greater consistency across pharmacogenomic studies. Here we provide specific responses to the main results reported by Mpindi *et al.*¹

The FIMM and CCLE studies used a similar experimental protocol that included the CellTiter-Glo pharmacological assay, as opposed to the Syto60 assay used in CGP. As pointed out by the authors¹, there were also parts of the experimental protocols that were different between all the three studies, effectively preventing perfect replication of *in vitro* molecular and pharmacological profiles. In our initial analysis², we showed that drug sensitivity measures for paclitaxel and lapatinib were more consistent between the GlaxoSmithKline (GSK) dataset and CCLE, than between GSK and CGP or CCLE and CGP. Given that both GSK and CCLE used the CellTiter-Glo assay, we concluded that the use of different pharmacological readouts has a substantial effect on the consistency of drug sensitivity measurements. The results from Mpindi *et al.*¹ further confirm this observation. However, a recent study from Genentech suggested that laboratory-specific effects might induce even greater biases than the use of different readouts⁶. Indeed, Haverty *et al.*⁶ showed that, although their CellTiter-Glo screen was more concordant

BRIEF COMMUNICATIONS ARISING

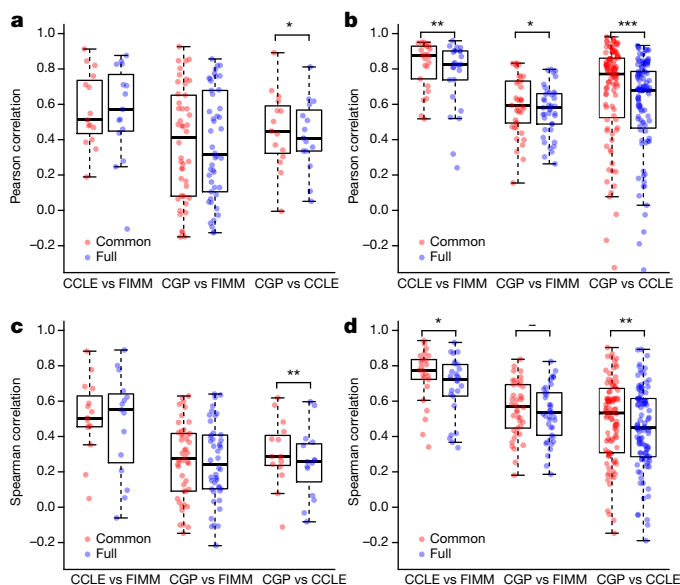


Figure 1 | Consistency between drug sensitivity measures (AUC) across FIMM, CGP and CCLE. a–d, AUC values were computed with the PharmacGx package using the full concentration range (blue) or the concentration range shared between two studies (red). a, b, Boxplots of Pearson correlations between AUC values across (a) and between (b) cell lines. c, d, Boxplots of Spearman rank correlations between AUC values across (c) and between (d) cell lines. The box and horizontal bar represent the interquartile range and median of the correlation coefficients, respectively, and the whiskers denote the most extreme data points. Significance of the difference between AUC correlations computed using the full or common concentration range is calculated using a one-sided Wilcoxon signed-rank test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; ‘—’ denotes $0.05 \leq P < 0.10$.

with CCLE, their new drug sensitivity data generated using the Syto60 assay were more consistent to their previous screen than CGP, despite the use of the same pharmacological assay⁶. It therefore remains unclear whether there are other experimental factors that drive the observed inconsistencies between large-scale pharmacogenomic studies and further argues for a detailed analysis of experimental protocols.

Similar to Pozdeyev *et al.*⁷ and the Comment by Bouhaddou *et al.*⁸, the authors investigated whether sensitivity metrics computed from the drug concentration range shared between studies yield higher consistency than the published metrics computed on the full (only partially overlapping) concentration range using different curve fitting algorithms. Concurring with previous results, Mpindi *et al.*¹ showed that the modified area under the curve (AUC) statistic (referred to as the drug sensitivity score (DSS)) computed on the shared concentration range (harmonized) was better correlated between studies than published AUC values (unharmonized). To test whether this marginal but significant improvement was due to the use of the same drug dose–response curve modelling or the choice of concentration range, we reproduced the authors’ analysis using our PharmacGx package⁹ and used the same curve-fitting algorithm for the FIMM CGP and CCLE studies. We observed significantly higher correlations, across and between cell lines, for AUC values computed using a shared concentration range for the CGP and CCLE comparison ($P < 0.05$, Wilcoxon signed-rank test; Fig. 1). Although this observation held true for the correlations between cell lines for all comparisons, restricting AUC computation to the common concentration range did not yield significantly higher correlation across cell lines for CCLE versus FIMM and CGP versus FIMM (Fig. 1). Our results confirm that restricting the analysis to the common concentration range improves consistency

between CGP and CCLE, and to a lesser extent with the FIMM dataset.

In our original report², we computed correlation coefficients for each individual drug across cell lines, which was relevant to the overall goal of the CCLE and CGP studies to discover new genomic biomarkers of drug responses in order to increase the emergence of ‘personalized’ treatment regimens. Mpindi *et al.*¹ also computed the correlation of cell line sensitivity data across drugs (referred to as between cell lines in their presentation; see Supplementary Information). While the application of the across cell line correlation analysis is more relevant for identifying biomarkers predictive of response to individual drugs, we agree with Mpindi *et al.*¹ that correlations across and between cell lines should be compared in a consistent manner, as was done in our published in-depth reanalysis of CCLE and CGP¹⁰. Consistent with Mpindi *et al.*¹, our results clearly demonstrate that the overall correlation across cell lines is lower than the correlation between cell lines (Supplementary Fig. 1). However, gene expression data are significantly more concordant between studies than the drug response summary statistics (half-maximum inhibitory concentration (IC_{50}) and AUC) values in all comparisons ($P < 0.002$, Wilcoxon rank-sum test). Consequently, our observation that gene expression data are significantly more correlated than pharmacological response still holds.

As we argued previously², we agree with Mpindi *et al.*¹ that there is a need for harmonization of experimental protocols and cross-validation of large pharmacogenomic studies, and that doing so will improve robustness and reproducibility of the associated data.

Author A. C. Jin was a student in A.H.B.’s laboratory and left shortly after publication of the initial study, and did not participate in the writing of this Reply. Authors Z.S., P.S. and M.F. developed the PharmacGx software package, which enabled the analyses presented here; A.G. helped with the comparison of the different drug sensitivity metrics, and participated in the interpretation of the results and writing of this Reply.

Methods

The methods are described in detail in our published reanalysis¹⁰ and in the Supplementary Information. The code and associated files required to reproduce this analysis are publicly available on the cdrug-rebuttals GitHub repository (<https://github.com/bhklab/cdrug-rebuttals>). The procedure to set up the software environment and run our analysis pipeline is provided in the Supplementary Information. This work complies with the guidelines proposed previously¹¹ in terms of code availability and replicability of results.

Zhaleh Safikhani^{1,2}, Nehme El-Hachem³, Petr Smirnov¹, Mark Freeman¹, Anna Goldenberg^{4,5}, Nicolai J. Birkbak^{6,7}, Andrew H. Beck^{8,9}, Hugo J. W. L. Aerts^{9,10,11}, John Quackenbush^{10,12} & Benjamin Haibe-Kains^{1,2,5,13}

¹Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario M5G 2M9, Canada.

email: bhaibeka@uhnresearch.ca

²Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada.

³Institut de recherches cliniques de Montréal, Montreal, Quebec H2W 1R7, Canada.

⁴Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada.

⁵Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada.

⁶The Francis Crick Institute, London NW1 1AT, UK.

⁷University College London Cancer Institute, London WC1E 6BT, UK.

⁸Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA.

⁹Harvard Medical School, Boston, Massachusetts 02115, USA.

¹⁰Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.

BRIEF COMMUNICATIONS ARISING

¹¹Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

¹²Harvard TH Chan School of Public Health, Boston, Massachusetts 02115, USA.

¹³Ontario Institute of Cancer Research, Toronto, Ontario M5G 1L7, Canada.

1. Mpindi, J. P. *et al.* Consistency in drug response profiling. *Nature* **540**, <http://dx.doi.org/10.1038/nature20171> (2016).
2. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).
3. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
4. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
5. Pemovska, T. *et al.* Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. *Cancer Discov.* **3**, 1416–1429 (2013).
6. Haverty, P. M. *et al.* Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* **533**, 333–337 (2016).
7. Pozdeyev, N. *et al.* Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. *Oncotarget* <http://dx.doi.org/10.18632/oncotarget.10010> (2016).
8. Bouhaddou, M. *et al.* Drug response consistency in CCLE and CGP. *Nature* **540**, <http://dx.doi.org/10.1038/nature20580> (2016).
9. Smirnov, P. *et al.* PharmacGx: an R package for analysis of large pharmacogenomic datasets. *Bioinformatics* **32**, 1244–1246 (2016).
10. Safikhani, Z. *et al.* Revisiting inconsistency in large pharmacogenomic studies. *F1000Research* <http://dx.doi.org/10.12688/f1000research.9611.1> (2016).
11. Sandve, G. K., Nekrutenko, A., Taylor, J. & Hovig, E. Ten simple rules for reproducible computational research. *PLoS Comput. Biol.* **9**, e1003285 (2013).

Supplementary Information is available in the online version of the paper.

doi:10.1038/nature20172

Drug response consistency in CCLE and CGP

ARISING FROM B. Haibe-Kains *et al. Nature* **504**, 389–393 (2013); doi:10.1038/nature12831

The Cancer Cell Line Encyclopedia¹ (CCLE) and Cancer Genome Project² (CGP) are two independent large-scale efforts to characterize genomes, mRNA expression, and anti-cancer drug dose-responses across cell lines, providing a public resource relating cellular biochemical context to drug sensitivity. A recent study³ analysed correlations between reported dose-response metrics and found inconsistency between CCLE and CGP, thus questioning the validity of not only these, but also other current and future costly large-scale studies. Here, we examine two

metrics of drug responsiveness (slope and area under the curve) that we derive from the original CCLE and CGP data, and find reasonable and statistically significant consistency. Our results revive confidence that the CCLE and CGP drug dose-response data are of sufficient quality for meaningful analyses. There is a Reply to this Comment by Safikhani, Z. *et al. Nature* **540**, <http://dx.doi.org/10.1038/nature20581> (2016).

CCLE and CGP share 2,520 dose-responses across 285 cell lines and 15 drugs, but cells were treated with different dose ranges. To compare

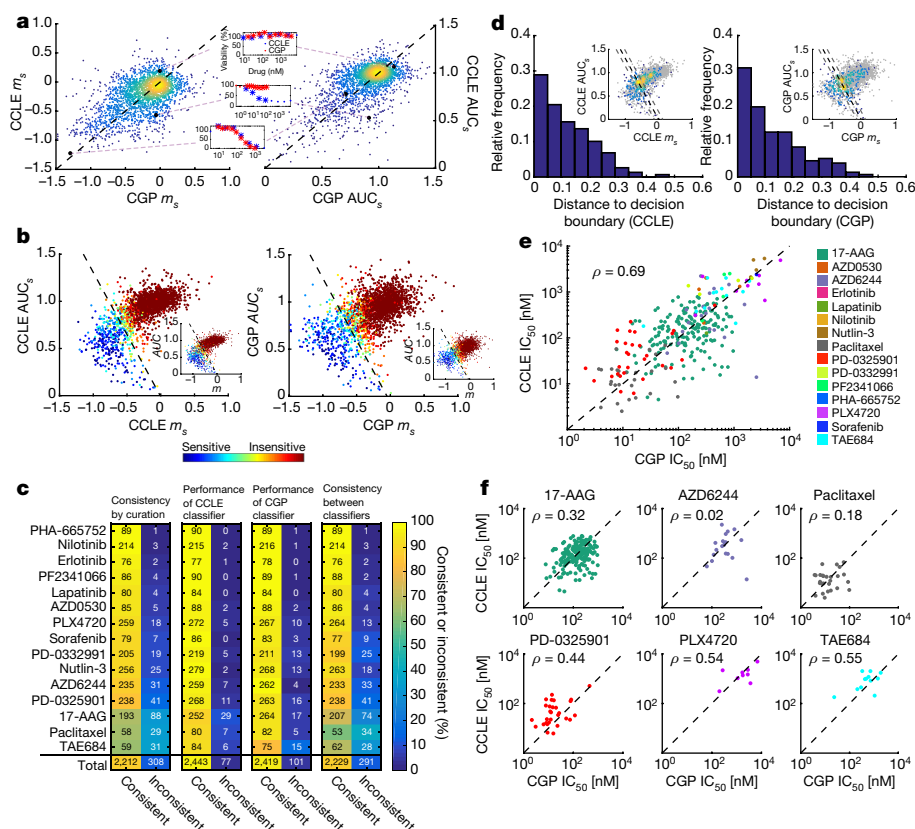


Figure 1 | Consistency between pharmacological data in CCLE and CGP. **a**, Slope (m_s ; left) or area under the curve (AUC_s ; right) of the dose-response curves for all overlapping drug/cell line pairs (2,520) in CCLE and CGP, considering only the shared dose range (denoted by subscript s). All m_s and AUC_s values were normalized based on the respective drug dose range, to facilitate comparison across drugs (see Supplementary Methods). Colour indicates density of dots. The black dashed line is $x = y$. In example dose-response curves, stars represent the shared dose range. **b**, Relationship between m_s and AUC_s for each database (inset m and AUC defined with the entire dose range as opposed to the shared dose range). The SVM classifier decision boundary divides the plot into sensitive and insensitive drug/cell line pairs, as indicated by the black dashed line. Slope and y -intercept of boundary line for CCLE: $m = -1.32$, $b = -0.01$; CGP: $m = -1.31$, $b = -0.06$. Colour of dots indicates the mean of the binary classifications from eight manual curators; blue indicates a unanimous sensitivity rating, green a very uncertain rating, and red a unanimous insensitivity rating. **c**, Consistency (left) and inconsistency (right) of classification methods broken down by drug. Far left plot shows manual curation consistency between CCLE and CGP. Middle left plot shows consistency between the manual curation data from CCLE and the CCLE

SVM classifier. Middle right plot shows consistency between the manual curation data from CGP and the CGP SVM classifier. Far right plot shows consistency between the CCLE SVM classifier used to classify CGP data and the CGP SVM classifier used to classify CCLE data. Colour indicates percentage consistency as denoted by the colour bar. Numbers denote number of observations, black for consistent, white for inconsistent. **d**, Inconsistent drug/cell line pairs based on manual curation results. Histograms bin the Euclidian distance between each discrepantly classified drug/cell line pair (that is, called sensitive in one database and insensitive in the other) and the decision boundary (black dashed line) in the AUC_s versus m_s plots for CCLE (left) or CGP (right). In inset, coloured dots indicate drug/cell line pairs that were classified discrepantly in CCLE and CGP. Colour corresponds to density of dots. Black dashed line indicates the decision boundary for the SVM classifier. Grey dashed lines indicate a Euclidian distance of 0.1 from the decision boundary in either direction. **e**, IC_{50} values from all sensitive cell line/drug combinations as determined by SVM classifier in CCLE or CGP. The black dashed line is $x = y$. **f**, IC_{50} values from all sensitive cell line/drug pairs (same as in Fig. 1e) stratified by drug, for drugs having at least 5 points. All correlation coefficients are Pearson.

CCLE and CGP dose–responses, we calculated a common viability metric (0–100%) across a shared \log_{10} -dose range, and computed slope (m_s) and area under the curve (AUC_s) values (in which subscript ‘s’ denotes the shared dose range) (Fig. 1a). This analysis revealed surprisingly good quantitative agreement between the two studies (m_s : population Pearson correlation coefficient (ρ) = 0.52, $P < 10^{-16}$; AUC_s : ρ = 0.61, $P < 10^{-16}$). Furthermore, since a small m_s or large AUC_s value indicates insensitivity, these data suggest that most cell lines are insensitive to the majority of tested drugs (~85%, Fig. 1a, b). Characterizing such insensitive trends with a sigmoid model meant for sensitive cell lines (that is, half-maximum inhibitory concentration, IC_{50}) may lead to incorrect dataset consistency conclusions.

To evaluate consistency of sensitivity classification between the two studies, we first asked eight people to curate binary sensitivity manually (all dose–response curves and their manually curated classification results are provided in the Supplementary Information). For manual curation, only data from a single database within the shared dose range was presented on each plot, and the order of plot presentation was randomized with respect to the study, the drug, and the cell line for each curator (see Extended Data Figs 1 and 2). Using the manual curation results, we built a separate support vector machine (SVM) classifier for each study with m_s and AUC_s as predictors (Fig. 1b). Both SVMs performed well (Fig. 1c, middle two plots), and the decision boundaries are independently similar for CCLE and CGP (Fig. 1b, black dashed line). These SVM classifiers also seem to parse data derived from the full (not shared) range of drug doses effectively (Fig. 1b; insets; m and AUC without subscript s), which may be important for future, database-specific analyses.

The manual curation data along with the SVM classifiers allowed evaluation of consistency between CCLE and CGP in terms of binary sensitivity classification (Fig. 1c). Comparison of manual curation results shows high (~88%) and statistically significant consistency between the two studies overall (Cohen’s kappa (κ) = 0.53 ± 0.025), and for most individual drugs (Fig. 1c, far left). Using the CCLE SVM to classify CGP data, and vice versa (Fig. 1c, far right), also yielded high and statistically significant consistency (88%, κ = 0.55 ± 0.025). These results strongly suggest that drug dose–response data in the CCLE and CGP can be considered consistent when used to classify binary sensitivity.

The drugs 17-AAG, paclitaxel and TAE684 account for 48% of the inconsistent drug/cell line pairs. We hypothesized that most of these and other inconsistent drug/cell line pairs would be located near the SVM decision boundary. The primary reason is because this boundary necessarily travels through the region of AUC_s – m_s space where determining binary sensitivity is the most challenging for manual curators (Fig. 1b, cyan to yellow dots denote uncertainty among curators). If true, then this would imply that a main factor driving the observed inconsistency is self-induced: imposing a strict cutoff. Indeed, most such inconsistent points are located close to the decision boundary; for CCLE 53% of the inconsistent points are within 0.1 distance units from the decision boundary, and 51% for CGP (Fig. 1d). Manual inspection of these inconsistent binary classification cases also supports this interpretation (Supplementary Data 1). We do observe some strongly inconsistent drug cell/line pairs (for example, Fig. 1a inset-middle, and Supplementary Data 1), but these are relatively rare, and are highly likely to be located far from a decision boundary. These results suggest that inconsistency between the two studies on the level of binary classification is, to a large extent, a result of the information loss associated with collapsing a two-dimensional continuous description of drug sensitivity onto a single binary variable. Thus, we propose that drug sensitivity is better described as a spectrum (AUC and m) than as a binary classification.

We next re-calculated and compared IC_{50} data only from drug/cell line pairs determined to be sensitive in either CCLE or CGP by the SVM classifier (another requirement was the existence of a non-extrapolated IC_{50} value). We found good correlation between the two studies overall (Fig. 1e; ρ = 0.69, $P < 0.0001$). However, stratification by drug generally yields poor IC_{50} correlations (Fig. 1f). Thus, caution should be taken for inference of IC_{50} values for specific cell line/drug combinations from CCLE and CGP, despite consistency on the level of slope, area under the curve, and binary sensitivity classification. Haibe-Kains *et al.*³ stratified IC_{50} by drug for sensitive and insensitive lines (IC_{50} values for insensitive lines are unreliable), which contributed to their conclusion of inconsistency.

We conclude that the drug dose–response data in CCLE and CGP are acceptably consistent for most cases. Furthermore, we made no attempts to remove potentially suspect dose–response data, but doing so in future efforts could further facilitate data usability. That the two studies are this consistent is quite remarkable, given the different viability assays used, as well as inescapable confounding factors such as cell confluency, clonal variations, genomic drift, different drug suppliers/batches, laboratories/equipment and serum composition. This suggests that the measured genomic and gene expression parameters may provide a robust cellular context that dictates drug sensitivity.

Methods

For each drug/cell line pair found in both CCLE and CGP, we calculated the slope and AUC of each dose–response curve (percentage cell viability versus \log_{10} drug dose) only in the shared dose range. These values were normalized to account for different dose ranges used by each drug. One CCLE point and one CGP point defined boundaries of the shared dose range to maximize data coverage. IC_{50} values were calculated as the drug concentration needed to reach 50% cell viability (using a fit to a sigmoid response model) if within the shared dose range (see Supplementary Methods). All scripts and data needed to reproduce the figures, including the MATLAB code, are provided in Supplementary Data 2.

Mehdi Bouhaddou¹, Matthew S. DiStefano¹, Eric A. Riesel¹, Emilce Carrasco¹, Hadassa Y. Holzapfel¹, DeAnalisa C. Jones¹, Gregory R. Smith¹, Alan D. Stern¹, Sulaiman S. Somani¹, T. Victoria Thompson¹ & Marc R. Birtwistle^{1,2,3}

¹Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

email: marc.birtwistle@mssm.edu

²DTXs LINCS Center, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

³Systems Biology Center New York, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

Received 15 November 2014; accepted 13 October 2016.

1. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
2. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
3. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).

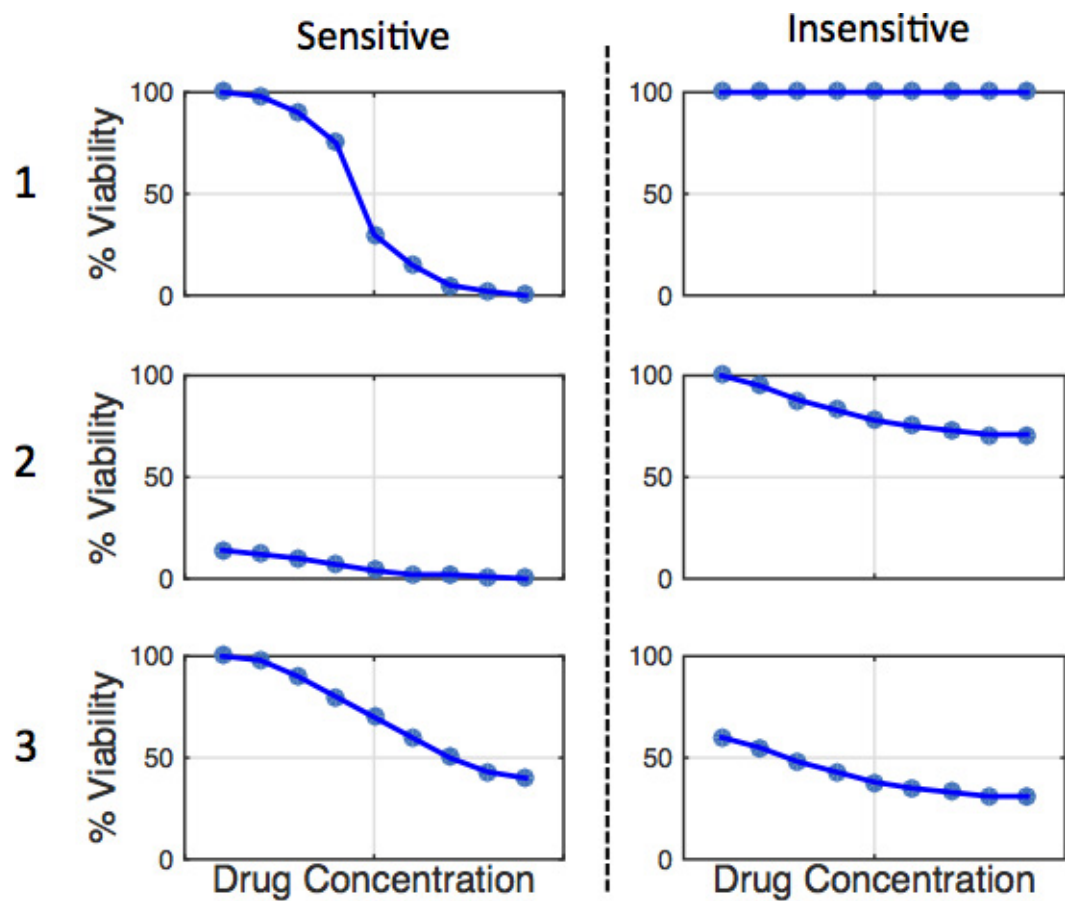
Supplementary Information is available in the online version of the paper.

Author Contributions M.R.B. conceived of the study. M.B. and M.R.B. performed the analyses and wrote the paper. M.S.D. prepared CCLE and CGP data for analysis and performed preliminary analyses. E.A.R., E.C., H.Y.H., D.C.J., G.R.S., A.D.S., S.S.S. and T.V.T. served as manual curators.

Competing Financial Interests Declared none.

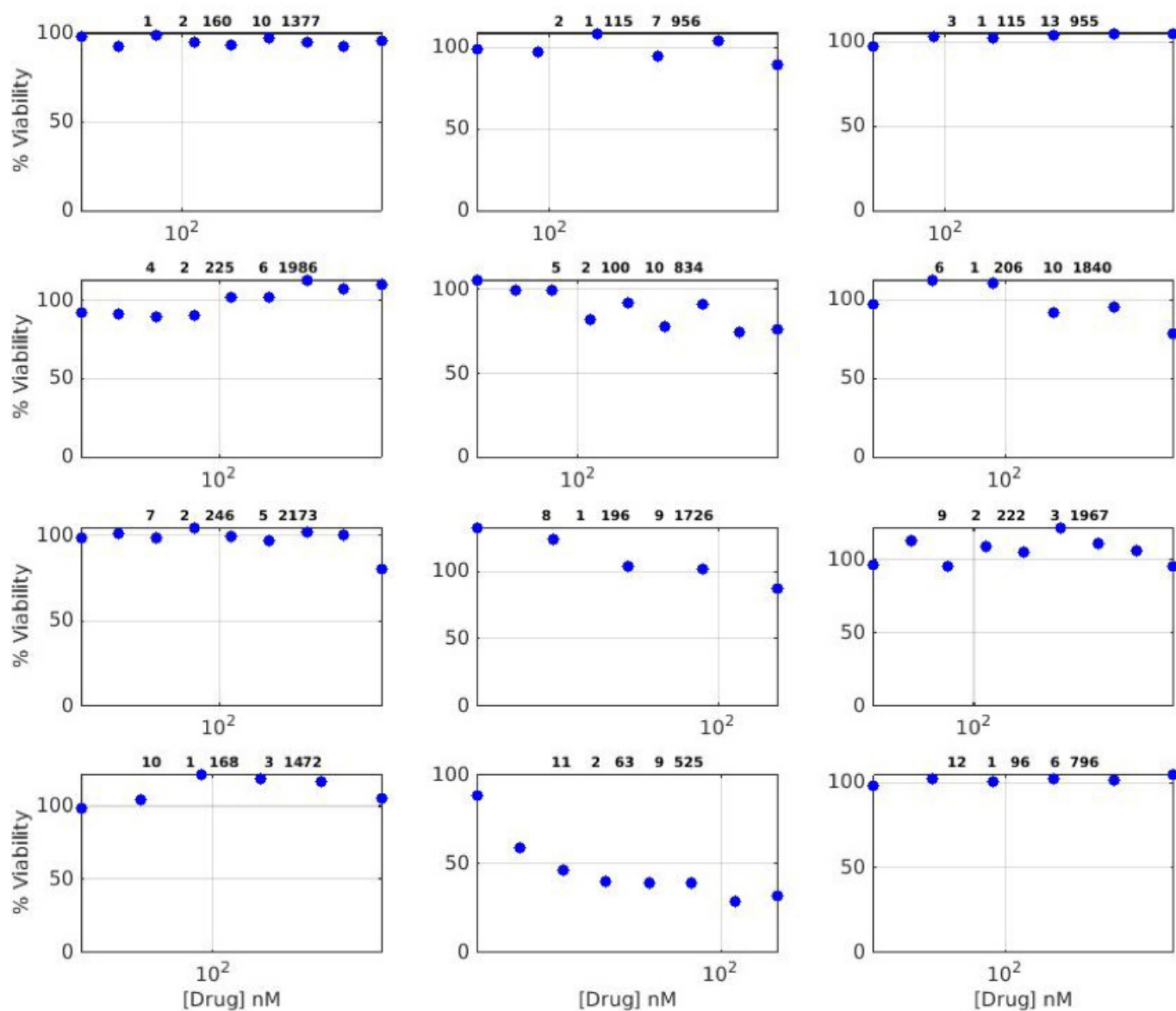
doi:10.1038/nature20580

BRIEF COMMUNICATIONS ARISING



Extended Data Figure 1 | Examples of typical sensitive versus insensitive dose–response curves. This document was given to manual curators as example dose–response curves. These idealized data represent various dose–response curves one might encounter in CCLE and/or CGP and indicate how they should be classified.

BRIEF COMMUNICATIONS ARISING



Extended Data Figure 2 | Examples of what the manual curators received. This is one page, as an example, from the data given to manual curators, which they were instructed to rate as either sensitive or insensitive.

Safikhani *et al.* replyREPLYING TO M. Bouhaddou *et al.* *Nature* 540, <http://dx.doi.org/10.1038/nature20580> (2016)

In the accompanying Comment¹, the authors make two main claims: (1) that viability metrics computed over the drug concentration range shared between datasets yield higher consistency than the same metrics computed over the full, often only partially overlapping, drug concentration range; and (2) that binary drug sensitivity classification (insensitive versus sensitive) as determined by manual curators increases the consistency of pharmacological profiles between the Cancer Genome Project (CGP)² and Cancer Cell Line Encyclopedia (CCLE)³. We appreciate the innovative approach followed by the authors, and our reanalysis confirms the marginal, but statistically significant, increase in consistency achieved through the use of viability metrics computed across a reduced but common range. However, our results indicate that manual classification of drug dose–response curves does not significantly increase the agreement between drug sensitivity calls compared to computational approaches. Notably, it is unclear whether the authors' manual approach will improve reproducibility of the biomarker discovery process, as collapsing of complex curves into discrete categories may result in a substantial information loss. Here we provide specific responses to the main results reported by Bouhaddou *et al.*¹

Similar to Pozdnyev *et al.*⁴ and the Comment by Mpindi *et al.*⁵, the authors investigated whether sensitivity metrics computed from the drug concentration range shared between CGP and CCLE yield higher consistency¹. Using the authors' code¹, we were able to implement their slope (m_s) and area under the curve (AUC_s) metrics (in which subscript 's' denotes shared dose range) in our PharmacGx platform⁶ with a minor improvement of the m_s metric to prevent highly sensitive cell lines with flat drug dose–response curves to be classified as insensitive (see Supplementary Methods). We compared the AUC and m metrics computed from the full and shared drug concentration range for the pooled set of drug sensitivities. Our implementation of the drug dose–response curve fitting and sensitivity computation further improved the authors' results: initial correlation for m_s ($\rho = 0.52$) and AUC_s ($\rho = 0.61$) both increased to 0.67. We then tested whether the common viability metrics constituted significant improvement over computations on the full drug concentration range. We observed a small but statistically significant improvement for both the m_s and AUC_s metrics (test of difference in correlations, $P < 0.01$; Supplementary Fig. 1). Stratifying our analysis per drug, we observed that the improvement in consistency, although significant, was marginal, with the exception of nilotinib (Supplementary Fig. 2). However, most of the drugs still yielded poor consistency ($\rho < 0.5$), which is in line with both our initial report⁷ and our more recent reanalysis⁸.

The authors investigated discretization of their continuous metrics to test whether binary classification would yield higher consistency, as estimated by the overall percentage agreement in drug sensitivity calls. However, such a statistic does not take into account the agreement that would be expected purely by chance owing to the large proportion of cell lines being insensitive to the tested drugs. The Matthews correlation coefficient (MCC)⁹ addresses this issue. It is a balanced measure that can be used when the classes are of different sizes, and its significance can be computed using the χ^2 statistic for binary classes (Supplementary Methods). We illustrate the case of four drugs with different patterns of consistency in Supplementary Fig. 3. Although all four drugs yield an overall agreement of greater than 92%, they exhibit a wide range of MCC values. Nilotinib is a good example of a consistent drug phenotype across cell lines (MCC = 0.86; Supplementary Fig. 3a).

PLX4720 yields moderate consistency (MCC = 0.68; Supplementary Fig. 3b). AZD0530 and erlotinib show only poor consistency (MCC = 0.42 and –0.05, respectively; Supplementary Fig. 3c, d). These examples support MCC as an appropriate statistic to discriminate between highly consistent drug sensitivity calls and those with poor concordance. We therefore used the MCC to compare different classification schemes, including those proposed by the authors.

Recognizing the difficulty of summarizing drug dose–response curves computationally, Bouhaddou *et al.*¹ used an unconventional approach to increase the consistency of drug sensitivity calls: they gathered a team of eight curators and asked them to classify each drug dose–response curve as either sensitive or insensitive. The authors report a Cohen's kappa (κ) value of 0.53, which is in line with our estimated MCC value of 0.53 (Supplementary Fig. 4). The authors qualified their manual classification as a high and statistically significant consistency. We disagree with the authors' claim¹ that their results provide evidence for high consistency. We refer them to the standards for strength of agreement for κ defined previously¹⁰, which would only classify observed consistency as moderate. More importantly, when classifications are stratified by drug, we do not observe a significant improvement of manual curation over the computational classifications based on AUC_s and m_s values ($P > 0.12$, Wilcoxon signed rank test; Supplementary Fig. 5). Consistent with our previous report, 10 out of 15 drugs (66.7%) yielded poor consistency (MCC < 0.5).

By pooling drug sensitivity data across drugs, the authors noticed a good quantitative agreement between the two studies, with estimated Pearson correlation coefficients (ρ) of 0.52 and 0.61 for AUC_s and m_s values, respectively¹; our improved implementation of their method increased the correlation to 0.67. Nevertheless, we disagree that this level of correlation constitutes evidence for good agreement, and define it as only moderate consistency based on the interpretation scale of our initial study⁷. More importantly, the common viability metrics only marginally improved consistency at the level of individual drugs (except for nilotinib), with most of the drug yielding inconsistent drug sensitivity values ($\rho < 0.5$; Supplementary Fig. 2). The authors made a similar observation¹ in their figure 1f, undermining their claim of drug response consistency in CGP and CCLE. Moreover, the main goal of CGP and CCLE consisted of finding new associations between molecular features and sensitivity to specific drugs^{2,3}. Since biomarkers are to be found for each drug separately, it is vital that pharmacological profiles are highly consistent at the level of individual drugs and not merely when averaged across a larger dataset.

In conclusion, our re-analysis of the new AUC_s and m_s metrics described by Bouhaddou *et al.*¹ showed that they represent a statistically significant improvement over the published drug sensitivity values, but the increase in consistency is only marginal for the vast majority of the drugs tested both in CCLE and CGP. Furthermore, manual classification of the drug dose–response curves does not appear to substantially improve the consistency of binary sensitivity calls over computational approaches and is not a scalable method. However, the authors¹ showed that manually classified drug dose–response curves could be used as a benchmark to train nonlinear computational predictors that could take into account the peculiar features of each individual dataset. Although there is no evidence that the authors' approaches¹ improve reproducibility of biomarker discovery for individual drugs, their work may open a new avenue of research in pharmacogenomics. Manual curation and further exploration of new

large datasets such as CTRPv2 (ref. 11) and GDSC1000 (ref. 12)—containing approximately 395,000 and 225,000 individual curves, respectively—will present major challenges, but the investment in these large pharmacogenomic warrants such efforts.

Author A. C. Jin was a student in A.H.B.'s laboratory and left shortly after publication of the initial study, and did not participate in the writing of this Reply. Authors Z.S., P.S. and M.F. developed the PharmacGx software package, which enabled the analyses presented here; A.G. helped with the comparison of the different drug sensitivity metrics, and participated in the interpretation of the results and writing of this Reply.

Methods

The methods are described in detail in the Supplementary Information. The code and associated files required to reproduce this analysis are publicly available on the cdrug-rebuttals GitHub repository (<https://github.com/bhklab/cdrug-rebuttals>). The procedure to set up the software environment and run our analysis pipeline is provided in the Supplementary Information. This work complies with the guidelines proposed previously¹³ in terms of code availability and replicability of results.

Zhaleh Safikhani^{1,2}, Nehme El-Hachem³, Petr Smirnov¹, Mark Freeman¹, Anna Goldenberg^{4,5}, Nicolai J. Birkbak⁶, Andrew H. Beck^{7,8}, Hugo J. W. L. Aerts^{8,9,10}, John Quackenbush^{9,11} & Benjamin Haibe-Kains^{1,2,5,12}

¹Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario M5G 2M9, Canada.

email: benjamin.haibe.kains@utoronto.ca

²Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada.

³Institut de recherches cliniques de Montréal, Montreal, Quebec H2W 1R7, Canada.

⁴Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada.

⁵Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada.

⁶The Francis Crick Institute, University College London, London NW1 1AT, UK.

⁷Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA.

⁸Harvard Medical School, Boston, Massachusetts 02115, USA.

⁹Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.

¹⁰Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

¹¹Harvard School of Public Health, Boston, Massachusetts 02115, USA.

¹²Ontario Institute of Cancer Research, Toronto, Ontario M5G 1L7, Canada.

1. Bouhaddou, M. *et al.* Drug response consistency in CCLE and CGP. *Nature* **540**, <http://dx.doi.org/10.1038/nature20580> (2016).
2. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
3. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
4. Pozdeyev, N. *et al.* Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. *Oncotarget* <http://dx.doi.org/10.18632/oncotarget.10010> (2016).
5. Mpindi, J. P. *et al.* Consistency in drug response profiling. *Nature* **540**, <http://dx.doi.org/10.1038/nature20171> (2016).
6. Smirnov, P. *et al.* PharmacGx: An R package for analysis of large pharmacogenomic datasets. *Bioinformatics* **32**, 1244–1246 (2015).
7. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).
8. Safikhani, Z. *et al.* Revisiting inconsistency in large pharmacogenomic studies. *F1000Research* <http://dx.doi.org/10.12688/f1000research.9611.1> (2016).
9. Matthews, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **405**, 442–451 (1975).
10. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).
11. Seashore-Ludlow, B. *et al.* Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.* **5**, 1210–1223 (2015).
12. Iorio, F. *et al.* A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754 (2016).
13. Sandve, G. K., Nekrutenko, A., Taylor, J. & Hovig, E. Ten simple rules for reproducible computational research. *PLoS Comput. Biol.* **9**, e1003285 (2013).

doi:10.1038/nature20581

PLANETARY SCIENCE

Pluto's telltale heart

Studies of a large frost-filled basin on Pluto show that this feature altered the dwarf planet's spin axis, driving tectonic activity on its surface, and hint at the presence of a subsurface ocean. **SEE LETTERS P.86, P.90, P.94 & P.97**

AMY C. BARR

In a culture of selfies, it's difficult to imagine a more compelling picture of Pluto than the image returned by NASA's New Horizons space probe in July 2015 (Fig. 1). Pluto put its best face forward for New Horizons, showing us an area of bright, smooth terrain, in the shape of a heart. A renewed love of Pluto, which was stripped of its planet-hood a decade ago, swelled in the hearts of planetary scientists and school-children alike. Four papers published in this issue of *Nature*^{1–4} show that the heart formed as a result of the interplay of slow deposition of frozen noxious chemicals, bitterly cold winds, cracking icy crusts, cryogenic buried oceans and planetary cartwheels. Pluto may be cute, but this is planetary science, after all.

Pluto is one of the dwarf planets in the Kuiper belt, a family of objects beyond the orbit of Neptune that are about half the size of Earth's Moon, and is composed of solid water ice and rock. On Pluto, an outer shell of water ice behaves like bedrock — the ice is shaped into mountains, fractures and faults by internal tectonic forces⁵, and forms impact craters when objects strike the dwarf planet⁶.

For decades since its discovery, astronomers have watched Pluto as it crossed the line of sight between Earth and distant stars. By measuring how the spectrum of the starlight changed during these events, researchers showed that Pluto has an atmosphere composed of nitrogen, methane and carbon monoxide⁷, all of which can be in the solid or the gaseous state at Plutonian temperatures. Pluto's axial tilt of 120° and eccentric orbit mean that the amount and pattern of sunlight that falls on its surface change dramatically over the course of its 90,560-day year. This variability drives volatile ices to condense in some places and sublime (be directly converted from solid to gas) in others — these ices therefore migrate across Pluto's surface.

The left half of the heart-shaped region on Pluto is informally known as Sputnik Planitia and is unlike any other geological feature in the Solar System. It consists of a depression

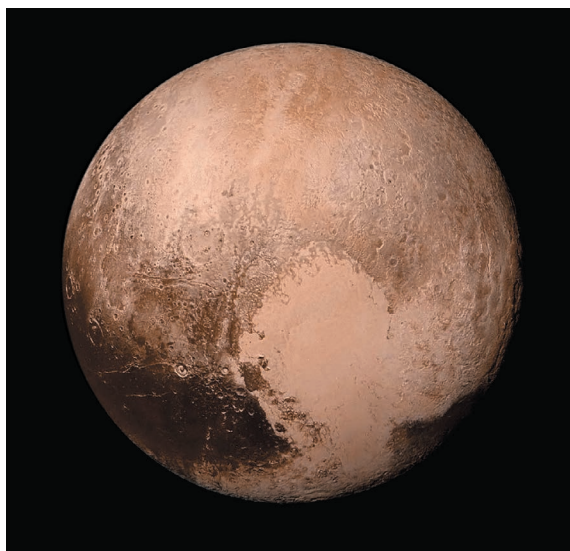


Figure 1 | Sputnik Planitia. This global view of Pluto was obtained by combining four high-resolution images from New Horizons' Long Range Reconnaissance Imager (LORRI) with colour data from its Ralph telescope. Four papers^{1–4} study Sputnik Planitia, the left half of the heart-shaped region of bright terrain that is visible in the centre of the image.

in Pluto's water-ice shell that is filled with the same ices that comprise the dwarf planet's atmosphere⁸. This deposit of atmospheric ices is about 4 kilometres thick^{7,9}, which is similar to the average depth of Earth's oceans. Its surface is smooth and only 10 million years old⁶. The four new studies use data from New Horizons to explain the origins of Sputnik Planitia and its effect on Pluto.

Bertrand and Forget¹ (page 86) simulate how nitrogen, methane and carbon monoxide frosts sublime from regions that are warmed by sunlight, and condense in cold or low-lying regions. Their simulations reproduce the time-evolution of global frost deposits on Pluto, recorded by decades of telescopic observations⁷, and show that a low-latitude basin could easily accumulate frost that is many kilometres thick. Hamilton and colleagues² (page 97) point out that the formation of a single frost deposit on Pluto is inevitable — as the frost begins to accumulate, that region reflects more sunlight and becomes colder, driving further deposition. Because this frost deposit is denser than the surrounding water ice, it creates its own depression.

Hamilton *et al.*, Keane *et al.*³ (page 90) and Nimmo *et al.*⁴ (page 94) describe how Pluto's interior would have reacted to the frost deposit. Pluto is not perfectly spherical — the gravity of its moon, Charon, causes Pluto to be egg-shaped. Therefore, in the minimum-energy configuration of the Pluto–Charon system, the long axes of Pluto and Charon are aligned. Charon always orbits above the same spot on Pluto, as though the two bodies were joined by a rigid stick.

However, when frost accumulated in Sputnik Planitia, the system was no longer in a minimum-energy configuration. Pluto rolled over, and its ice shell was fractured by the resulting tension and compression forces, creating canyons and mountains³. The line that now joins Pluto and Charon pierces the centre of Sputnik Planitia, a configuration that is favoured only if Sputnik Planitia represents a region of excess mass, despite the fact that it is a depression. Nimmo and colleagues show that such a mass excess is possible if the ice shell is thin and underlain by a liquid ocean, which is consistent with previous findings⁸.

These four studies are a testament to the deductive powers of modern planetary science. On the basis of a single set of images from New Horizons, the authors have used lessons learnt from analysing other planets to unravel the mystery of Pluto's heart. With no future Pluto missions planned, are these studies verifiable? Bertrand and Forget predict a disappearance of frost in Pluto's northern hemisphere in the coming decades, which could be observed using telescopes. But the other studies offer few suggestions, apart from numerical modelling, to test their hypotheses.

However, the processes that were responsible for the formation of Sputnik Planitia also shaped other planetary bodies. For example, a similar frost migration and deposition occurred on Mars¹⁰. And the Moon¹¹, Mars¹² and Saturn's moon Enceladus¹³ have undergone reorientation due to a loading of material on their crusts. Perhaps, as simulations of these processes improve through further

comparisons between models and data, our understanding of Pluto will be enhanced, providing support or opposition for these interpretations. ■

Amy C. Barr is at the Planetary Science Institute, Tucson, Arizona 85719-2395, USA. e-mail: amy@psi.edu

CELL BIOLOGY

Double agents for mitochondrial division

Mitochondrial organelles — the energy powerhouses of the cell — must divide and fuse dynamically to function. It emerges that two distinct dynamin enzymes enable mitochondrial division. SEE LETTER P.139

HEIDI M. MCBRIDE & ADAM FROST

To paraphrase the biologist François Jacob, the dream of every mitochondrion — the cell's energy-producing organelle — is to become two mitochondria. On page 139, Lee *et al.*¹ report that two different dynamin proteins act sequentially in the process of membrane division that divides a mitochondrion into two. One aids the initial mitochondrial-membrane constriction, and the other enables the final stages of lipid constriction necessary for membrane fission. Mitochondria have an essential role in cellular function, and gaining a better understanding of how they divide might shed light on human diseases caused by their dysfunction.

Mitochondria exist as a highly dynamic, interconnected network of organelle structures known as a reticulum, which undergoes repeated rounds of fusion and division². These remodelling events distribute mitochondria throughout the cell, aid the maintenance of mitochondrial numbers and metabolic capacity, and ensure that mitochondrial genomes are efficiently dispersed throughout the reticulum¹.

Cell biologists have long sought to understand how membrane-bound organelles divide, and how they pinch off membranous vesicles to transport proteins and nutrients throughout the cell. Research in this area has often focused on the dynamin family of proteins, GTPase enzymes that assemble in a collar-like structure around the constricting lipid 'necks' of budding membrane-bound vesicles or tubular organelle structures^{2,3}. The hydrolysis of the molecule GTP provides energy that drives changes in dynamin structure; these conformational changes squeeze dynamin-encircled lipid membranes together until membrane division occurs,

1. Bertrand, T. & Forget, F. *Nature* **540**, 86–89 (2016).
2. Hamilton, D. P. *et al.* *Nature* **540**, 97–99 (2016).
3. Keane, J. T., Matsuyama, I., Kamata, S. & Steckloff, J. K. *Nature* **540**, 90–93 (2016).
4. Nimmo, F. *et al.* *Nature* **540**, 94–96 (2016).
5. Hammond, N. P., Barr, A. C. & Parmentier, E. M. *Geophys. Res. Lett.* **43**, 6775–6782 (2016).
6. Moore, J. M. *et al.* *Science* **351**, 1284–1293 (2016).
7. Elliot, J. L. *et al.* *Astron. J.* **134**, 1–13 (2007).
8. Grundy, W. M. *et al.* *Science* **351**, aad9189 (2016).

9. Stern, S. A. *et al.* *Science* **350**, aad1815 (2015).
10. James, P. B., Kieffer, H. H. & Paige, D. A. in *Mars* (eds Matthews, M. S., Kieffer, H. H., Jakosky, B. M. & Snyder, C.) 934–968 (Univ. Arizona Press, 1992).
11. Keane, J. T. & Matsuyama, I. *Geophys. Res. Lett.* **41**, 6610–6619 (2014).
12. Perron, J. T., Mitrovica, J. X., Manga, M., Matsuyama, I. & Richards, M. A. *Nature* **447**, 840–843 (2007).
13. Nimmo, F. & Pappalardo, R. T. *Nature* **441**, 614–616 (2006).

mitochondrial tubule to the point of fission was unknown.

Lee *et al.* noticed that cells lacking dynamin 2 had an excess of fused mitochondria, which provided a hint that dynamin 2 might have a role in mitochondrial division. Using high-resolution light microscopy, the authors monitored dynamin 2 at sites of mitochondrial division. Simultaneously observing dynamin 2 and Drp1 revealed how these proteins act to fully constrict a mitochondrial tubule. Using electron microscopy, Lee and colleagues measured the size of mitochondrial constrictions that accumulated in the absence of dynamin 2. This revealed that Drp1 mediates mitochondrial membrane constriction to a diameter of approximately 100 nm, and dynamin 2 then completes the membrane-tubule constriction to the point of fission (Fig. 1).

This finding was a surprise because dynamin 2 had previously been associated mainly with the fission of vesicles that mediate endocytosis, the process of uptake of extracellular material³. Another surprise was the sequential use of two different dynamins at separate stages of organelle-membrane constriction, which hadn't been observed before.

Drp1 and dynamin 2 share some evolutionarily conserved features with the rest of the dynamin family, particularly in the structural domains that drive assembly of these proteins around a constriction. Each dynamin also binds specific proteins and lipids, which determine where it functions in the cell. The recruitment of Drp1 to the mitochondria occurs through binding to mitochondrial receptor proteins⁴, including Mif, MiD49 and MiD51. This recruitment is tightly controlled

through a process known as fission³.

Some dynamins² can constrict vesicles with diameters of 100 nanometres down to the approximately 10-nm diameter required for membrane fission. However, mitochondria have diameters ranging from 500 to 1,000 nm (ref. 2), and two lipid membranes. This mitochondrial scale means that substantially greater remodelling is needed for mitochondrial fission than for the membrane fission of vesicles.

The sites on the mitochondrial-membrane surface at which division will ultimately occur are initially marked by contact with the endoplasmic reticulum (ER), another cellular organelle². These contacts are thought to initiate mitochondrial-membrane constriction by polymerizing actin-protein filaments, which provides force for remodelling of the mitochondrial tubule². The mitochondrial dynamin-related protein Drp1 is recruited to these pre-constricted contact sites and assembles in helical and ring-like structures on the membrane surface². However, whether Drp1 assembly and enzyme activity constrict the

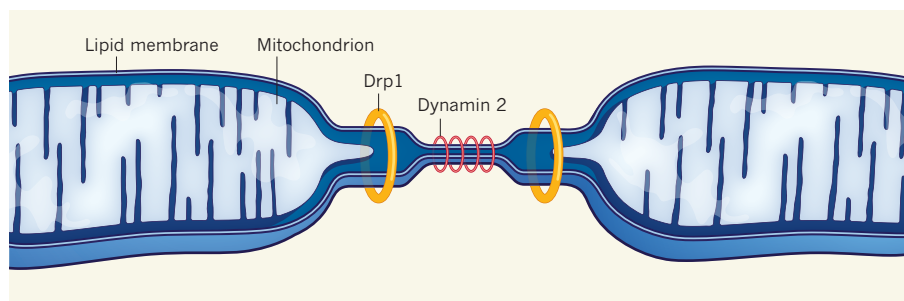


Figure 1 | Different roles for dynamins in mitochondrial division. When a replicating mitochondrial organelle divides in two, its lipid membrane constricts in the middle until the lipid layers fuse and separate. The dynamin enzyme Drp1 binds to the mitochondrial membrane and partially constricts it. Lee *et al.*¹ have found that another dynamin, known as dynamin 2, binds and constricts the mitochondrial membrane further to enable the final step of lipid fusion and organelle division.

through modifications to Drp1 that alter its conformation and activity, providing a link from cell signalling and metabolism to mitochondrial dynamics².

A central question raised by Lee and colleagues' work is how dynamin 2 is recruited to mitochondria. This answer might already be close at hand, because mitochondrial isoforms of key dynamin 2 binding proteins are already known. These include the BAR-domain protein endophilin B1 (ref. 5) and the phosphatase enzyme synaptonemin 2A (ref. 6), which targets the lipid phosphoinositide. However, the function of these proteins in mitochondrial division has not been well established. Another protein worth investigating is the dynamin 2 binding partner sorting nexin 9 (SNX9), which regulates formation of vesicles from mitochondria⁷.

A key area for future work will be determining how the sequential constriction of mitochondria proceeds through the three stages described by Lee and colleagues: from the initial ER-induced constrictions to the Drp1 assembly that constricts the diameter of the mitochondrion, and finally the regulated recruitment of dynamin 2 to the Drp1-constricted site that is required for membrane division. These are separable events, which should allow molecular dissection of each step individually.

For example, microscopy by Lee and colleagues revealed that, in the absence of dynamin 2, Drp1 was recruited to a constricted mitochondrial tubule, giving rise to a 'frustrated' division event in which a narrow mitochondrial tubule covered with Drp1 pulled and contracted, but did not divide. Earlier work⁸ showed that loss of Drp1 led to ER-mediated partial constrictions, creating mitochondrial structures similar to beads on a string. Studying such intermediate states will help researchers to understand how each of the two dynamins is recruited and assembled to divide mitochondria.

Lee and colleagues' discovery means that mitochondria should be taken into consideration when interpreting observations of dynamin inhibition by genetic or pharmacological means. For example, mutations in dynamin 2 are linked to human disorders such as Charcot-Marie-Tooth disease and centronuclear myopathy⁹. Although research has focused on errors in endocytosis and vesicle transport to explain these diseases⁹, alterations in mitochondrial dynamics might contribute to the pathology. Similar neurodegenerative characteristics are also caused by inherited mutations in mitochondrial proteins, including Drp1 (ref. 10). By demonstrating that dynamin 2 has functions in mitochondrial division, Lee and colleagues open up avenues of investigation into how mitochondrial dynamics affect disease. ■

Heidi M. McBride is in the Department of Neurology and Neurosurgery, Montreal

Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada.

Adam Frost is in the Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94143, USA.

e-mails: heidi.mcbride@mcgill.ca; adam.frost@ucsf.edu

1. Lee, J. E., Westrate, L. M., Wu, H., Page, C. & Voeltz, G. K. *Nature* **540**, 139–143 (2016).
2. Labbé, K., Murley, A. & Nunnari, J. *Annu. Rev. Cell Dev. Biol.* **30**, 357–391 (2014).
3. Antonny, B. et al. *EMBO J.* **35**, 2270–2284 (2016).

4. Osellame, L. D. et al. *J. Cell Sci.* **129**, 2170–2181 (2016).
5. Karbowski, M., Jeong, S. Y. & Youle, R. J. *J. Cell Biol.* **166**, 1027–1039 (2004).
6. Nemoto, Y. & De Ca74.millii, P. *EMBO J.* **18**, 2991–3006 (1999).
7. Matheoud, D. et al. *Cell* **166**, 314–327 (2016).
8. Friedman, J. R. et al. *Science* **334**, 358–362 (2011).
9. González-Jamett, A. M. *J. Neurochem.* **128**, 210–223 (2014).
10. Waterham, H. R. et al. *N. Engl. J. Med.* **356**, 1736–1741 (2007).

This article was published online on 23 November 2016.

For a related paper on mitochondrial dynamics, see page 74.

QUANTUM COMPUTING

Efficient fault tolerance

Dealing with errors in a quantum computer typically requires complex programming and many additional quantum bits. A technique for controlling errors has been proposed that alleviates both of these problems.

DANIEL GOTTESMAN

Quantum computers have great potential for solving certain computational problems — such as simulating chemical reactions or exotic quantum materials — that seem to be intractable for regular computers. However, although the leading experiments have achieved error rates of less than one error in every 1,000 operations¹, a large quantum algorithm might require millions of operations, and even a single error could result in the wrong answer. Future large-scale quantum computers will therefore need to be programmed in a 'fault-tolerant' way so that they can get the correct answer in spite of a low but non-negligible error rate. Unfortunately, existing error-correction approaches come with a major drawback: they require many additional quantum bits (qubits). Writing in *Physical Review X*, Yoder et al.² propose a more efficient method for performing certain fault-tolerant quantum operations

using fewer qubits than existing approaches.

To deal with errors in a quantum computer, the computer's memory is protected by a quantum error-correcting code. The qubits that we wish to protect are supplemented by additional qubits that spread out the quantum information; a carefully designed code identifies errors by detecting changes in the correlations between the qubits. Errors in a quantum computation can be thought of as a disease that strikes certain qubits at random. The error-correction procedure is a cure for the disease, but one that is in limited supply — if the error rate is low enough, only a few qubits get sick and they can be cured efficiently, but if there are too many errors, our treatment abilities will be overwhelmed.

As long as the qubits simply sit there, error correction is relatively straightforward. But quantum computations require interactions between the qubits. If one qubit is sick, it can infect other qubits with which it interacts, and the error can propagate. Therefore, the

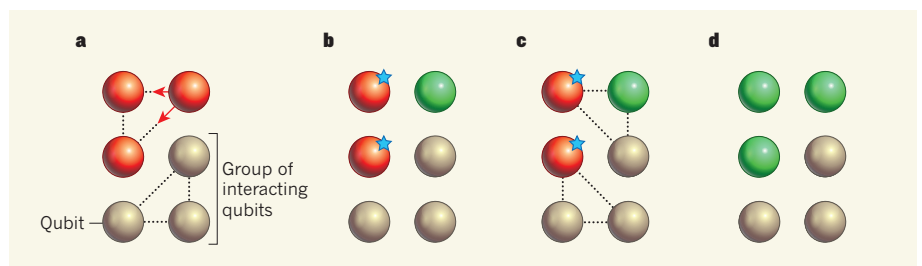


Figure 1 | Outbreak prevention. Yoder et al.² propose a technique for eliminating errors in a quantum computer. **a**, The authors first allow the quantum bits (qubits) of the computer to interact in small groups. An error (red) that occurs in one qubit can spread to other members of the same group (indicated by the arrows). **b**, After a potential outbreak, the qubits are immediately checked and the source of the error is identified. This source is corrected (green), and qubits that were in the same group are marked for tracking and later correction (denoted by stars). **c**, The qubits then interact in different groups. **d**, Finally, the qubits are checked again and the previously marked qubits corrected.

error-correcting code described above needs to be upgraded to a fault-tolerant protocol, which allows quantum computations to be performed while keeping the qubits protected against errors.

A quantum computation can be broken down into quantum gates; these are the most elementary computational steps, such as adding two qubits together. The most straightforward approach to creating fault-tolerant quantum gates is called a transversal gate. For such gates, the qubits are quarantined into small groups, and interactions are allowed only within a group. Therefore, errors can be treated before they spread into an epidemic.

Unfortunately, a full quantum computation cannot use only transversal gates³ — another kind of fault-tolerant gate is always needed. The predominant strategy to overcome this limitation is to prepare special quantum states known as magic states⁴, which are ‘injected’ into the quantum computation to perform the desired gate. This injection procedure can be carried out with a controllable spread of errors, but reliably making a magic state is difficult.

One solution is to make many magic states, some of which will be tainted with errors. Using a procedure called magic-state distillation⁵, the states are compared with one another to find a few that are safe to inject into the computation. As one might expect, this procedure is extremely inefficient, requiring many attempts at making magic states to obtain even a small number that are almost error-free. It has been estimated⁶ that more than 90% of the resources in a large-scale quantum computer would be devoted to magic-state distillation.

Consequently, there have been many efforts in the past five years either to reduce the resources needed for magic-state distillation or to find alternatives. One approach has been to devise more-efficient procedures for the distillation. There has been some success in this direction⁷, but the process remains resource-intensive, perhaps because we lack a perfect understanding of how and why many magic-state distillation protocols work. Various research groups^{8–10} have also proposed alternative ways of performing non-transversal fault-tolerant gates, the latest being Yoder and colleagues.

In their work, the authors propose slightly relaxing the quarantine that is used in transversal gates and allowing the qubits to interact in groups of limited size (Fig. 1). Of course, this can lead to a small outbreak in which errors propagate within a group of qubits. However, by keeping track of each group and promptly performing error correction after a potential outbreak, the spread of errors can be limited to make sure that it remains under control. Yoder *et al.* find that their approach uses half as many qubits as does magic-state distillation — even if failed attempts at distillation are neglected.

But Yoder and collaborators only compare the efficiency of a quantum error-correcting

code that is mediocre at eliminating errors, and which is therefore not a leading candidate for fault tolerance. The authors show how their construction can be applied to other codes, but not the most advanced ones being studied today^{11,12}. In particular, their protocol involves tracking outbreaks of errors that stem from a single qubit. Future work that uses the most advanced codes will require many outbreaks to be controlled simultaneously, which will probably call for a more complicated procedure. Nevertheless, building a large-scale, fault-tolerant quantum computer is such a daunting task that every new approach is welcome. ■

Daniel Gottesman is at the Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada.
e-mail: dgottesman@perimeterinstitute.ca

CELL BIOLOGY

Sort of unexpected

To reach the cell surface, membrane proteins are first targeted to an organelle called the endoplasmic reticulum. Several targeting pathways are known, but it now emerges that there is yet another pathway. [SEE LETTER P.134](#)

MARTIN R. POOL

Secreted proteins and membrane proteins are made and transported to the cell's exterior through what is known as the secretory pathway: a set of linked membrane-bound compartments that includes an organelle called the endoplasmic reticulum (ER). Protein sorting to the ER has been studied for more than 40 years¹, and three pathways^{2–4} that transport proteins to the ER have been identified. However, there is still plenty to learn about this process because, on page 134, Aviram *et al.*⁵ report a previously unknown pathway that can target membrane proteins to the ER in the yeast *Saccharomyces cerevisiae*.

A key step in a protein's entry to the secretory pathway is the initial sorting process that directs proteins to the ER. The first ER-targeting mechanism to be identified involved signal sequences² — stretches of 6–15 hydrophobic amino-acid residues⁶ present at the amino terminus of a protein. Signal sequences are used to direct a protein to the ER, where the sequences are cleaved off. The transmembrane domains of membrane proteins can also function as non-cleaved signal sequences.

Signal sequences are recognized and bound by an evolutionarily conserved⁷ RNA-protein complex called the signal recognition particle (SRP). SRP is found in the cytoplasm and on the surface of the ER. It also binds to ribosomes^{2,3}, the cellular RNA-protein complexes that translate messenger RNA molecules to

1. Gaebler, J. P. *et al.* *Phys. Rev. Lett.* **117**, 060505 (2016).
2. Yoder, T. J., Takagi, R. & Chuang, I. L. *Phys. Rev. X* **6**, 031039 (2016).
3. Eastin, B. & Knill, E. *Phys. Rev. Lett.* **102**, 110502 (2009).
4. Shor, P. W. *Proc. 37th Annu. Symp. Found. Comput. Sci.* 56–65 (IEEE, 1996).
5. Bravyi, S. & Kitaev, A. *Phys. Rev. A* **71**, 022316 (2005).
6. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. *Phys. Rev. A* **86**, 032324 (2012).
7. Bravyi, S. & Haah, J. *Phys. Rev. A* **86**, 052329 (2012).
8. Paetznick, A. & Reichardt, B. W. *Phys. Rev. Lett.* **111**, 090505 (2013).
9. Jochym-O'Connor, T. & Laflamme, R. *Phys. Rev. Lett.* **112**, 010505 (2014).
10. Hill, C. D., Fowler, A. G., Wang, D. S. & Hollenberg, L. C. L. *Quant. Inf. Comput.* **13**, 439–451 (2013).
11. Raussendorf, R. & Harrington, J. *Phys. Rev. Lett.* **98**, 190504 (2007).
12. Gottesman, D. *Quant. Inf. Comput.* **14**, 1338–1371 (2014).

This article was published online on 16 November 2016.

synthesize proteins. When SRP recognizes a signal sequence on a protein undergoing translation³, it targets this protein and the associated ribosomal complex to the ER, where protein translation is completed and the protein is simultaneously transferred into the ER^{2,3}.

However, not all signal sequences are recognized by SRP. In yeast, signal sequences with less hydrophobic cores use an SRP-independent targeting pathway⁴ mediated by the ER-membrane protein Sec62. Unlike the SRP system, which recognizes proteins only while they are being translated, proteins can undergo post-translational targeting to the ER through the Sec62 pathway. The Sec62 pathway also operates in mammalian cells^{8,9} and is the most common ER-targeting pathway for secretory proteins shorter than around 100 amino acids^{8,9}, probably because such proteins are released from the ribosome before they can efficiently engage with SRP.

Membrane proteins are targeted to the ER by the SRP-targeting pathway. However, tail-anchored membrane proteins (those that have their transmembrane domain at the carboxy terminus of the protein) are released from the ribosome before SRP can engage with them^{10,11}. These membrane proteins are targeted to the ER by a different route, through the action of guided entry of tail-anchored (GET) proteins^{10,11}. Although many tail-anchored membrane proteins fulfil essential functions, yeast GET mutants are not lethal¹². A lack of GET proteins has only modest effects on the cell,

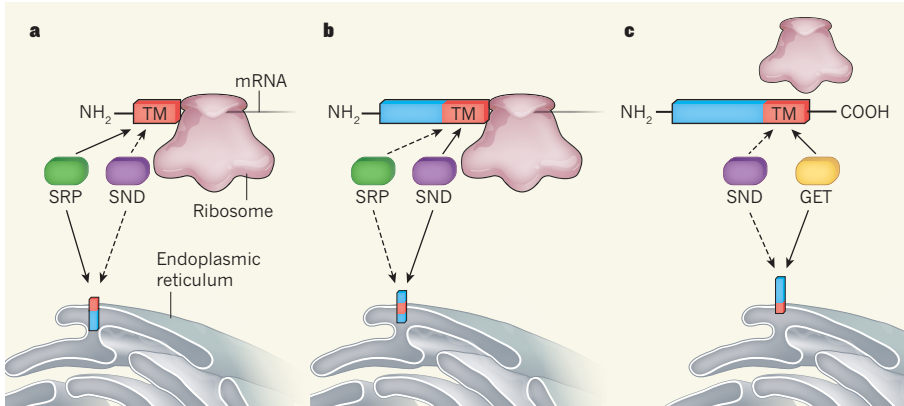


Figure 1 | Ways to target membrane proteins to the secretory pathway. Proteins reach the cell exterior after moving through the secretory pathway, which comprises a series of membrane-bound compartments. The first step in the transport of membrane proteins through the secretory pathway occurs when these proteins are targeted to an organelle called the endoplasmic reticulum (ER). **a, b,** If a membrane protein contains a transmembrane (TM) domain at the amino terminus (**a**) or in the middle of the protein (**b**), this TM domain is recognized by the SRP protein during translation of messenger RNA by the ribosomal protein-synthesis complex. SRP binds the TM domain and targets the protein and ribosomal complex to the ER. **c,** If a membrane protein contains a TM domain at its carboxy terminus, the protein is bound by a GET protein after translation and targeted to the ER. Aviram *et al.*⁵ identified a previously unknown pathway that uses SND proteins to target membrane proteins that have a TM domain in the middle of the protein (**b**) to the ER. Filled and dashed arrows indicate the proposed main and secondary pathways, respectively, for targeting a protein to the ER depending on the location of the TM domain in the protein structure. Membrane protein shown in blue, apart from the TM domain which is red.

hinting that other ER-targeting pathways for membrane proteins must compensate when the GET machinery is absent.

Aviram and colleagues investigated ER-targeting mechanisms using the yeast Gas1 membrane protein, which moves to the ER in a manner that is SRP-independent and only partially dependent on the GET and Sec62 proteins¹³. The authors used a microscopy technique to monitor the cellular location of Gas1 in some 6,000 yeast strains that have different genetic mutations to identify genes associated with defective ER targeting. As expected, Aviram *et al.* found that GET or Sec62 mutations resulted in ER-targeting defects and led to mis-targeted Gas1 aggregates in the cytoplasm.

The authors identified three other mutant yeast strains that also formed Gas1 aggregates owing to defective ER targeting. These mutants did not affect the ER targeting of a protein that is usually sorted through the SRP-dependent pathway, and Aviram *et al.* named the ER-targeting components that they identified as SRP-independent targeting (SND) proteins. Snd1 encodes a cytoplasmic protein, and Snd2 and Snd3 encode membrane proteins of the ER.

Ribosomes translating ER-targeted proteins often become associated with the ER membrane. Therefore, to find which proteins are targeted by the SND pathway, Aviram and colleagues used a ribosomal-tagging technique¹⁴ in living cells that enabled them to isolate and purify ER-associated ribosomes and determine the nucleotide sequences of the associated mRNA molecules.

The authors compared the mRNA sequences isolated from wild-type and SND-mutant cells

to identify mRNA sequences that associate with the ribosome in an SND-dependent manner. Most of the SND-dependent mRNAs that they found encoded proteins that have a transmembrane domain located away from the N terminus of the protein. By contrast, mRNAs for proteins that had an N-terminal signal sequence or transmembrane domain were associated with ER-tagged ribosomes independently of SND.

Aviram *et al.* report that changing the placement of a transmembrane domain in an ER-targeted membrane protein had a strong effect on whether or not the protein underwent SND-dependent ER targeting (Fig. 1). The authors found that the SND machinery seems to target transmembrane sequences located away from the N-terminal region at the start of a protein sequence, whereas SRP is associated with targeting signal sequences and transmembrane domains at or near the N-terminal region of a protein.

Does the SND machinery preferentially interact with long polypeptide chains as they are being translated, or do short polypeptide chains preferentially bind SRP? Some ribosomal complexes bound to mRNA can recruit SRP before the signal sequence has been translated¹⁵. Gaining a better understanding of the interaction of SRP with the ribosome might provide some answers. One possibility is that SRP has preferential access to ribosomes with growing polypeptide chains while the chains are short, and that this preferential access is lost once the polypeptide chain extends, possibly because of the engagement of other proteins that might bind the ribosome or the polypeptide chain.

Many proteins seem to use more than one pathway to reach the ER, and such pathway compensation probably explains why the SND-dependent pathway remained undiscovered for so long. Aviram and colleagues observed that yeast cells with low SRP levels grow normally only if the SND machinery is present, and that inactivating the SND machinery in such a strain is lethal. Conversely, they found that overexpressing SND proteins reverses the severe growth defect that accompanies complete loss of SRP. These data are consistent with a model in which the SND pathway targets proteins with N-terminal transmembrane domains to the ER when the SRP pathway is not functional.

The SND- and GET-targeting pathways also seem to overlap, because Aviram *et al.* report that inactivating both pathways leads to severely compromised insertion of tail-anchor membrane protein into the ER and to lethality. Correct ER targeting of proteins is essential because membrane proteins can be toxic if they accumulate within the cell¹⁶. Having overlap in the ER-targeting pathways might provide added robustness by enabling cells to efficiently target membrane proteins under a wide range of physiological conditions and stresses.

Aviram and colleagues identified a related version of Snd2 in mammalian cells, and it will be interesting to learn whether an SND-dependent ER-targeting pathway operates in these cells. The authors' identification of the SND-targeting pathway adds an additional level of complexity to protein sorting in the secretory pathway, and opens up exciting research avenues for investigating the regulation of this pathway and its relationship to known ER-targeting mechanisms. ■

Martin R. Pool is in the Division of Molecular and Cellular Function, School of Biological Sciences, Faculty of Biology, Medicine & Health, University of Manchester, Manchester M13 9PT, UK.
e-mail: martin.r.pool@manchester.ac.uk

1. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 835–851 (1975).
2. Cross, B. C. S., Sinning, I., Lührink, J. & High, S. *Nature Rev. Mol. Cell Biol.* **10**, 255–264 (2009).
3. Zhang, X. & Shan, S. *Annu. Rev. Biophys.* **43**, 381–408 (2014).
4. Deshaies, R. J. & Schekman, R. *J. Cell Biol.* **109**, 2653–2664 (1989).
5. Aviram, N. *et al. Nature* **540**, 134–138 (2016).
6. von Heijne, G. *J. Mol. Biol.* **184**, 99–105 (1985).
7. Pool, M. R. *Mol. Membr. Biol.* **22**, 3–15 (2005).
8. Lakkaraju, A. K. K. *et al. Mol. Biol. Cell* **23**, 2712–2722 (2012).
9. Lang, S. *et al. J. Cell Sci.* **125**, 1958–1969 (2012).
10. Hegde, R. S. & Keenan, R. *J. Nature Rev. Mol. Cell Biol.* **12**, 787–798 (2011).
11. Denic, V. *Trends Biochem. Sci.* **37**, 411–417 (2012).
12. Schuldiner, M. *et al. Cell* **134**, 634–645 (2008).
13. Ng, D. T., Brown, J. D. & Walter, P. *J. Cell Biol.* **134**, 269–278 (1996).
14. Jan, C. H., Williams, C. C. & Weissman, J. S. *Science* **346**, 1257521 (2014).
15. Chartron, J. W., Hunt, K. C. L. & Frydman, J. *Nature* **536**, 224–228 (2016).
16. Rane, N. S. *et al. Dev. Cell* **15**, 359–370 (2008).

BIOGEOCHEMISTRY

Projections of the soil-carbon deficit

Changes in the amount of carbon stored in soil might be a crucial feedback to climate change. Experimental field studies show that warming-induced soil carbon losses are greatest where carbon stocks are largest. [SEE LETTER P.104](#)

ERIC A. DAVIDSON

Global soil carbon is a bit like the US federal deficit: the quantity of carbon is massive, its annual change is the difference between two large input and output terms (akin to revenues and expenditures), and its changes add up over decades to have major consequences for society. Changes in the planet's soil-carbon stocks result mostly from a modest difference between the main input term — plant carbon derived from photosynthesis — and the main output term, which is respiratory losses of carbon dioxide from microbial decomposition of soil organic matter. Both photosynthesis and respiration will respond to climate change, but predicting how these responses will affect the modest difference in photosynthetic input and respiratory output (and thus the effect on the storage of carbon in the soil) is challenging. Crowther *et al.*¹ have synthesized the results from 49 soil-warming experiments, and on page 104 they report that losses from carbon-rich Arctic soils might tip the global balance of soil-carbon storage towards a net positive feedback to climate change.

Unfortunately, we have few direct observations of the effects of warming on stocks of soil carbon, and projections of Earth-system models (ESMs) agree neither in the sign nor the magnitude of its change². Most ESMs simulate the rate of soil-carbon decomposition using simple temperature functions (typically, either the Q_{10} or Arrhenius functions), which are derived from short-term (days to years) laboratory and field observations. However, the sizeable fraction of carbon that normally resides in the soil for decades is of greatest relevance to climatic feedbacks, and the decomposition of such carbon may respond differently to changing climate than does decomposition observed during short-term studies³.

Crowther *et al.* analysed the results of 49 soil-warming experiments performed at sites across North America, Europe and Asia (Fig. 1). A few of these studies reported results after less than a year of experimental warming, but most were multi-year, and several lasted ten years or more. To normalize results obtained using different methodologies, the authors divided reported warming-induced

changes in soil carbon by the degree-years product — a metric obtained by multiplying the experimental treatment intensity (degrees warmed) by the duration of the experiment (number of years).

The authors found that the ratio of carbon lost per degree-year was largest when the study sites' soil-carbon stocks were large — something that occurs commonly at high latitudes. Warming can also increase photosynthesis and inputs of plant carbon to soil, but this apparently does not keep up with carbon losses in high-latitude ecosystems. When soil-carbon stocks are lower, which is often the case at lower latitudes, the responses of plant inputs and microbial decay seem roughly to even out, resulting in smaller observed changes in soil carbon (modest gains and losses).

Crowther *et al.* explored the significance of their findings across the world by applying a spatially distributed climate-change scenario to a global soil database using a statistical model that calculates carbon loss per degree-year as a function of the initial carbon stock. On average,

soils in high-latitude regions have large carbon stocks and undergo the greatest warming, and so such regions emerge as an amplified source of CO₂ to the atmosphere (Fig. 1).

An additional complication for projecting carbon losses, however, is that microbial processes in the soil seem to adjust to changes of temperature. This can be due to changes in the microbial communities, alterations in the optimal temperature for their enzyme functions, or exhaustion of readily decomposable organic carbon substrates³. Assuming that the thermal adjustment of soils to each increment of warming will happen within a year (which is the most conservative assumption with respect to carbon loss), the authors estimate a net loss of 55 petagrams of carbon (1 Pg is 10¹⁵ g) between 2015 and 2050 for a scenario in which global warming averages 2 °C; that carbon loss is equivalent to about 5 years of current anthropogenic emissions. They also show that losses could be much larger if thermal adjustments occur more slowly.

I believe that Crowther and colleagues' synthesis of results from soil-warming experiments is important because it provides independent constraints for decadal soil-carbon simulations based on experimental observations of degrees of warming, years of warming, initial carbon content and a suite of other factors that could affect the stability of carbon stored in the soil. Their new finding of a statistical dependence of decomposition on the initial carbon content of the soil may prove to be a useful benchmark for ESM simulations.

Soil-warming experiments do have limitations, however, including potential artefacts and costs that restrict their size, duration and

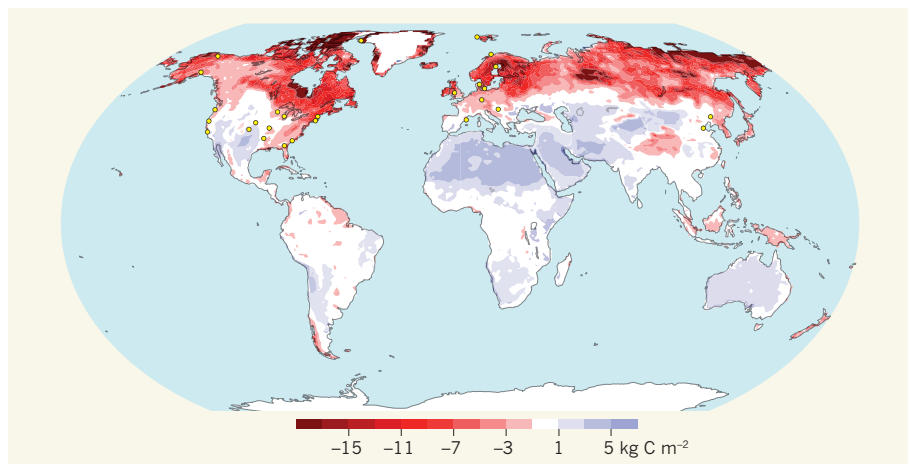


Figure 1 | Projections of the temperature vulnerability of soil carbon stocks. Crowther *et al.*¹ have analysed the results from 49 field studies in which soil was warmed at different sites around the world (sites are indicated with dots; some dots indicate general locations of more than one study) to assess how warming changes the density of soil carbon. They then extrapolated their findings using a statistical model to reveal the probable upper limit of changes in soil carbon by 2050 (shown in kilograms of carbon per square metre for the top 15 centimetres of soil), assuming a spatially distributed climate-change scenario in which average global temperature increases by 1 °C, and making certain assumptions about how long it takes for soil microbial processes to adapt to the increase. This analysis demonstrates that soils from high-latitude regions are especially vulnerable to net losses of carbon, which would act as a positive feedback to climate change. The extrapolations for tropical and desert regions are poorly constrained owing to lack of data from these regions. (Adapted from ref. 1.)

frequency. The studies analysed by Crowther *et al.* are still few and poorly distributed — for example, none has been performed in the Southern Hemisphere. Nevertheless, the new findings are highly informative because they consider experiments performed at different latitudes, and support long-held suspicions that the large soil-carbon stocks of the Arctic are vulnerable to decomposition in a warmer world^{4,5}.

This analysis does not explain why sites that have large carbon stocks tend to undergo the biggest losses. One explanation arises from the Arrhenius function, which predicts that 1 °C of warming should have a larger proportional effect on enzymatic activity at the low temperatures typical of high latitudes than it would at warmer, low latitudes. Kinetic theory

also indicates that the activation energies of enzymatic decomposition reactions (the minimum energy needed for such reactions to occur), and hence the temperature sensitivity of the reactions, should be high for the complex substrates that are common in humus-rich, wetland soils at high latitudes³.

But I suspect that an equally large or even larger factor is the exposure of increasing amounts of soil carbon to microbial decay when permafrost thaws and when flooded soils dry out. Models are only just beginning to include the effects of climate change on the exposure of soil carbon to microbial decomposition⁶, and Crowther *et al.* have provided a welcome empirical reference point for future simulations. Worryingly, their

findings plausibly suggest that soil carbon will be a source of a large biotic feedback that will accelerate climate change. ■

Eric A. Davidson is at the *Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, Maryland 21536, USA.*

e-mail: edavidson@umces.edu

1. Crowther, T. W. *et al.* *Nature* **540**, 104–108 (2016).
2. Todd-Brown, K. E. O. *et al.* *Biogeosciences* **11**, 2341–2356 (2014).
3. Davidson, E. A. & Janssens, I. A. *Nature* **440**, 165–173 (2006).
4. Billings, W. D. *Quat. Sci. Rev.* **6**, 165–177 (1987).
5. Abbott, B. W. *et al.* *Environ. Res. Lett.* **11**, 034014 (2016).
6. Wieder, W. R. *et al.* *Global Biogeochem. Cycles* **29**, 1782–1800 (2015).

(hundreds of kilometres per second) and produces narrow emission lines. For some AGNs, we observe both broad and narrow lines, whereas in others we see narrow lines only.

The theory of AGN unification^{4,5} posits that all AGNs are inherently the same — namely, these phenomena are accreting supermassive black holes that are surrounded by gas and dust in the shape of a torus. Therefore, if we are looking through this enshrouding material, the gas near the black hole will be blocked from our view and we will see narrow emission lines only. However, if we are looking through the opening of the torus, we will observe both the broad and the narrow lines.

Although this theory explains why the optical spectra of many AGNs do not show broad emission lines, ‘changing-look’ AGNs have been discovered in which the broad lines have either disappeared or appeared over time. Some of these phenomena can be explained through the lens of AGN unification — if

ASTRONOMY

A black hole changes its feeding habits

In the 1980s, the gas surrounding a black hole in a nearby galaxy began to emit much more radiation than before. This change has unexpectedly reversed in the past five years, questioning our understanding of these extreme phenomena.

STEPHANIE LAMASSA

Supermassive black holes are millions to billions of times more massive than the Sun and reside at the centres of almost all large galaxies such as the Milky Way¹. Some of these black holes are actively growing by accretion, a process in which material close to the black hole cannot escape the strong gravitational field and is ultimately pulled into the black hole. These growing black holes, or active galactic nuclei (AGNs), release large amounts of energy from the accretion process, making them visible across the Universe. Writing in *Astronomy & Astrophysics*, McElroy *et al.*² and Husemann *et al.*³ report that the emission spectrum of a nearby AGN has changed dramatically in the past five years, providing valuable insight into the physical processes involved in black-hole accretion.

Clues to the growth of black holes are imprinted in the optical spectra of AGNs. Just as a prism splits white light into its component colours, these spectra reveal the composition of matter, because atoms, ions and molecules emit light at specific wavelengths dictated by quantum physics. In the case of AGNs, light emitted by the accreting material energizes the surrounding gas, which produces prominent emission lines in the gas’s optical spectrum.

These emission lines also tell us about the motion of gas in the AGN. Because gas close to a black hole orbits at high velocities

(up to thousands of kilometres per second), the emission lines from these regions are broadened as a result of the Doppler effect: light emitted from gas moving away from Earth is shifted to longer wavelengths, whereas gas moving towards us emits light that is shifted to shorter wavelengths. Conversely, gas farther away from the black hole travels more slowly

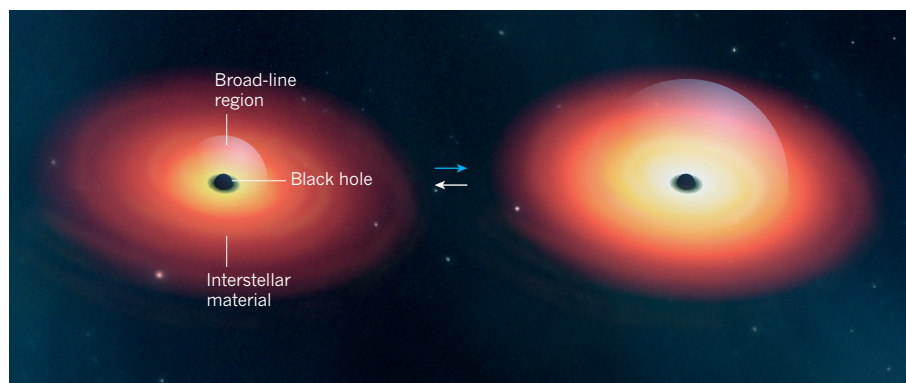


Figure 1 | The variability of Mrk 1018. Black holes can grow by pulling (accreting) interstellar material from their surroundings. This accretion process energizes the gas that surrounds the black hole, producing prominent lines in the gas’s emission spectrum. Because gas close to the black hole is extremely energetic, the emission lines from this region are broadened as a result of the Doppler effect. In the 1980s, prominent broad lines appeared⁹ in the optical spectrum of a black-hole system called Mrk 1018. The leading explanation for this change (indicated by the blue arrow) is that the accretion rate onto the black hole increased, energizing more of the gas and expanding the size of the broad-line region. McElroy *et al.*² and Husemann *et al.*³ have combined optical and X-ray observations to show that, in the past five years, Mrk 1018 has returned to its original state (indicated by the white arrow). The authors suggest that this transition is due to a decrease in the black-hole accretion rate.

the obscuring material is patchy, it can block or reveal the gas close to the black hole as it orbits⁶. However, variable obscuration cannot explain other changing-look AGNs^{7,8} such as Mrk 1018, the one studied by McElroy *et al.* and Husemann and colleagues.

In the 1980s, the optical spectrum of Mrk 1018 transitioned from having only narrow emission lines to also showing prominent broad lines⁹. Variable gas clouds were ruled out as being responsible for this transition. Instead, the favoured interpretation was that the gas was energized by an enhanced rate of accretion onto the black hole (Fig. 1). Such a change could have been caused by an increase in the gas supply near the black hole. For at least one changing-look AGN, it has been postulated that a nearby star was ripped apart by the gravitational (tidal) forces of the black hole, and that this provided the fuel to reignite the AGN¹⁰.

In 2015, Mrk 1018 was observed using the Multi Unit Spectroscopic Explorer (MUSE) installed on the Very Large Telescope in northern Chile, as part of a programme to study nearby AGNs. Using these observations, McElroy and colleagues serendipitously discovered that the optical spectrum of Mrk 1018 has changed again, returning to the state before the broad lines became prominent. By comparing the MUSE spectra with archival data from the 1980s to 2009, McElroy *et al.* show that this latest transition happened in the past five years. This result constrains the lifetime of the previous stage (in which the spectra showed broad lines) to between 25 and 30 years. Like the authors of the earlier analysis⁹, McElroy and collaborators suggest that the transition was probably caused by a change (in this case, a decrease) in the black-hole accretion rate, rather than by variable obscuration.

In a companion paper, Husemann and colleagues focus on the X-ray properties of Mrk 1018. In AGNs, X-rays are produced in a hot 'corona' that is closer to the black hole than the gas that emits broad emission lines, providing a direct probe of black-hole accretion. Husemann *et al.* compare data that were taken in 2016 by two space-based X-ray observatories — Chandra and the Nuclear Spectroscopic Telescope Array (NuSTAR) — with Chandra observations from 2010. The authors find that the amount of X-rays emitted by Mrk 1018 has significantly diminished over this time, and claim that this reduction is not due to absorption of the X-rays by the gas that surrounds the accreting material. Their results therefore support the conclusion that the transition of Mrk 1018 is not caused by variable obscuration.

From the hundreds of thousands of AGNs whose existence has been confirmed using optical spectroscopy, about 20 changing-look AGNs have been discovered, and only 3, including Mrk 1018, have had broad emission lines appear and then disappear^{7,11}. Many of the known changing-look AGNs were

discovered serendipitously, which has motivated dedicated searches to find more examples, taking advantage of spectroscopic data over multiple epochs of the Sloan Digital Sky Survey^{12–14}. The pace of discovery will increase in the coming decade, thanks to current and upcoming observatories such as Pan-STARRS and the Large Synoptic Survey Telescope, which will monitor the sky multiple times with the aim of detecting transient phenomena.

Although changing-look AGNs are rare, they are incredibly useful because they provide us with an opportunity to learn about the physics of accretion and the evolution of AGNs. The analyses of McElroy *et al.* and Husemann *et al.* teach us about the intermittency of black-hole accretion, suggesting that the picture is more complicated than the shutting down of an AGN. The timescale of these transitions also poses serious challenges to standard accretion theories¹⁵, which operate over periods that are seemingly too short^{8,10}. Changing-look AGNs therefore serve as valuable tools for refining theoretical models, using observations that were not previously possible. ■

BIOLOGICAL RHYTHMS

Wild times

Little is known about the biological rhythms that emerge from social behaviours in the wild. A study of shorebird pairs shows that rhythms of nest-incubation duties are mainly governed by strategies to avoid predators. [SEE LETTER P.109](#)

C. LOREN BUCK

Fundamental to all organisms, from bacteria to humans, is the daily oscillation of behaviour and physiology. During a 24-hour day, animals devote blocks of time to rest or activity¹. The pace of the daily rhythm is mainly governed by an organism's circadian clock, which is synchronized to the daily light-dark cycle^{2–5}. Biological rhythms can also be influenced by an organism's social environment; however, little is known about the factors that influence these social behavioural rhythms. On page 109, Bulla *et al.*⁶ report a large-scale study of the timing of daily nest-incubation duties shared between pairs of shorebirds.

Almost all studies of the daily rhythms of animals have used captive and individually housed laboratory strains under stringently controlled conditions. Such investigations have aided our understanding of biological rhythms; however, the survival of wild animals depends on their ability to shape their daily rhythms in response to complex external factors and, for many animals, on their ability to synchronize their activity with that of

Stephanie LaMassa is at NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. e-mail: stephanie.m.lamassa@nasa.gov

- Kormendy, J. & Ho, L. C. *Annu. Rev. Astron. Astrophys.* **51**, 511–563 (2013).
- McElroy, R. E. *et al. Astron. Astrophys.* **593**, L8 (2016).
- Husemann, B. *et al. Astron. Astrophys.* **593**, L9 (2016).
- Antonucci, R. *Annu. Rev. Astron. Astrophys.* **31**, 473–521 (1993).
- Urry, C. M. & Padovani, P. *Publ. Astron. Soc. Pacif.* **107**, 803–845 (1995).
- Storchi-Bergmann, T., Baldwin, J. A. & Wilson, A. S. *Astrophys. J.* **410**, L11 (1993).
- Denney, K. D. *et al. Astrophys. J.* **796**, 134 (2014).
- LaMassa, S. M. *et al. Astrophys. J.* **800**, 144 (2015).
- Cohen, R. D. *et al. Astrophys. J.* **311**, 135 (1986).
- Merloni, A. *et al. Mon. Not. R. Astron. Soc.* **452**, 69–87 (2015).
- Aretxaga, I., Joguet, B., Kunth, D., Melnick, J. & Terlevich, R. J. *Astrophys. J.* **519**, L123 (1999).
- Runnoe, J. C. *et al. Mon. Not. R. Astron. Soc.* **455**, 1691–1701 (2016).
- Ruan, J. J. *et al. Astrophys. J.* **826**, 188 (2016).
- MacLeod, C. L. *et al. Mon. Not. R. Astron. Soc.* **457**, 389–404 (2016).
- Shakura, N. I. & Sunyaev, R. A. *Mon. Not. R. Astron. Soc.* **175**, 613–632 (1976).

This article was published online on 23 November 2016.

other individuals. Although rhythms of social behaviour are evident in nature^{2,5,7,8}, there have been few investigations of the genetic and environmental factors that influence this phenomenon in wild animals.

Bulla and colleagues investigated a compelling natural model of social synchronization: the timing of shifts in nest attendance during egg incubation by individual pairs of shorebirds. Shorebird eggs require continuous incubation, and parents must precisely coordinate and synchronize their care of the nest. The authors report a 20-year study of incubation patterns of wild shorebirds, assembling a data set that includes 32 species, 729 pairs of birds and 91 populations that spanned high to low latitudes worldwide. They assessed incubation patterns using a combination of direct and indirect observation methods.

The authors found that the timing of individual incubation bouts ranged from just over 1 hour to an enormous 50 hours. The authors also calculated the durations of incubation periods, which reflect the average length of one cycle of shared parental incubation. Bulla *et al.* observed a diversity of incubation periods for the parental nest



Figure 1 | Nesting wild shorebirds use different strategies to avoid predators. Bulla *et al.*⁶ report a large-scale global study of nest-incubation timing patterns in pairs of wild shorebirds. **a**, The red knot (*Calidris canutus rogersi*) is camouflaged against its nesting background. The authors observed that this species exhibits long incubation bouts. This might help to avoid predation by reducing activity that could draw attention to the nest. **b**, The oystercatcher (*Haematopus ostralegus*) is not camouflaged against its nesting background, and instead actively responds to approaching predators. Bulla and colleagues noted that this species has short incubation bouts.

exchanges. Some pairs of birds synchronized their incubation periods around a 24-hour rhythm, whereas other pairs had incubation periods that were less than 20 hours or more than 28 hours. The authors also observed pairs of birds that exhibited free-running rhythms: near 24-hour rhythms that were not synchronized to the light–dark cycle. It is well established that most animals exhibit behavioural rhythms that are synchronized by the daily light–dark cycle, and these rhythms are known to be influenced by both genetic and environmental factors. Thus, the authors' observation of exceptional diversity in incubation rhythms, independent of habitat, both within the same species and across different shorebird species, was highly unexpected.

What factors might be responsible for the large diversity in incubation rhythms of wild shorebirds? Using sophisticated modelling approaches, the authors evaluated the contribution that genetics might make to this variability. Their results indicate that closely related bird species show similar incubation-rhythm patterns — a result consistent with other observations^{9–11} that biological rhythms are genetically heritable.

The authors investigated what other factors might shape incubation patterns. Nest incubation and food foraging are mutually exclusive behaviours, and it is reasonable to assume that incubation-bout length would be affected by a bird's energetic status. Small birds store less energy than large ones, use energy faster and should therefore need to forage more frequently. Thus, one might predict that bird species with a smaller body would exhibit more-frequent nest exchanges than larger species. However, Bulla *et al.* found no relationship between the body size of birds and the length of their incubation bouts.

Body-temperature regulation can also influence a bird's energetic demands. Birds in high-latitude, colder climates would be expected to deplete their energy stores more rapidly and show more-frequent incubation exchanges than birds in lower, warmer latitudes. However, Bulla and colleagues found

that shorebirds in cold climates had longer bouts of incubation than birds in warmer climates. Together, these results indicate that incubation-bout length is not driven by a bird's energetic status.

Bulla *et al.* identified one behavioural factor that significantly influenced the duration of incubation bouts: the strategy used by parents to reduce predator attacks on the nest. Shorebirds nest on the ground, and use either parental camouflage or direct, attacking or distracting strategies to deter predators from attacking their nests. The birds that avoid predation through camouflage have long incubation bouts. This approach lessens the parental changeover activity that could reveal a nest location to predators. By contrast, shorebirds that actively respond to intruding predators to dissuade them from approaching have short incubation bouts (Fig. 1).

The diversity observed in incubation rhythms of pairs of shorebirds, both within the same species and across different species, is impressive, but perhaps Bulla and colleagues' most striking finding is their observation that only 22% of the parental pairs adopted a 24-hour pattern for their incubation rhythms. This finding is intriguing, because synchrony with the environment is considered an advantageous characteristic¹², and such environmental asynchrony in animals is thought to have negative consequences, including associations with pathologies and early death^{13–15}. Humans with disrupted biological rhythms can develop symptoms that range from those seen in relatively benign jet lag, to those of more serious conditions, such as obesity¹⁶, Alzheimer's disease¹⁷ and cancer¹⁸.

The work by Bulla and colleagues reveals diversity, flexibility and plasticity in the timing of a social behaviour, and challenges the assumption that animals maintain synchrony with the 24-hour day–night cycle. Does being out of synchrony with the environment exact a physiological cost on shorebirds, and if so, is the cost equally distributed in a given pair of birds? The incubation rhythm will probably favour one parent over the other because

of differences in the natural phasing of food availability, predation risk and temperature for the parent away from the nest.

The finding that shorebird pairs of the same species can have substantially different incubation rhythms begs the question of how the rhythm is decided by a pair of birds. Is it a 'negotiation', or is the rhythm driven by the dominant partner? Is there year-to-year variation in the incubation rhythm for a given pair, and if so, what effect might fluctuating predator density have on the birds' rhythm? These and many other intriguing questions remain to be answered. ■

C. Loren Buck is in the Department of Biological Sciences and the Center for Bioengineering Innovation, Northern Arizona University, Flagstaff, Arizona 86011, USA.
e-mail: loren.buck@nau.edu

1. Dunlap, J. C., Loros, J. J. & DeCoursey, P. J. *Chronobiology: Biological Timekeeping* (Sinauer, 2009).
2. Davidson, A. J. & Menaker, M. *Curr. Opin. Neurobiol.* **13**, 765–769 (2003).
3. Mistlberger, R. E. & Skene, D. J. *Biol. Rev. Camb. Phil. Soc.* **79**, 533–556 (2004).
4. Favreau, A., Richard-Yris, M.-A., Bertin, A., Houdelier, C. & Lumineau, S. *Anim. Behav.* **77**, 983–989 (2009).
5. Castillo-Ruiz, A., Paul, M. J. & Schwartz, W. J. *Prog. Brain Res.* **199**, 267–280 (2012).
6. Bulla, M. *et al.* *Nature* **540**, 109–113 (2016).
7. Kronfeld-Schor, N., Bloch, G. & Schwartz, W. J. *Proc. Biol. Sci.* **280**, 20131354 (2013).
8. Bloch, G., Herzog, E. D., Levine, J. D. & Schwartz, W. J. *Proc. Biol. Sci.* **280**, 20130035 (2013).
9. Helm, B. & Visser, M. E. *Proc. R. Soc. Lond. B* **277**, 3335–3342 (2010).
10. Koskenvuo, M., Hublin, C., Partinen, M., Heikkilä, K. & Kaprio, J. *J. Sleep Res.* **16**, 156–162 (2007).
11. Mrosovsky, N. *Biol. Rev. Camb. Phil. Soc.* **71**, 343–372 (1996).
12. Yerushalmi, S. & Green, R. M. *Ecol. Lett.* **12**, 970–981 (2009).
13. Froy, O. *Physiology* **26**, 225–235 (2011).
14. Hurd, M. W. & Ralph, M. R. *J. Biol. Rhythms* **13**, 430–436 (1998).
15. Martino, T. A. *et al.* *Regul. Integr. Comp. Physiol.* **294**, R1675–R1683 (2008).
16. McEwen, B. S. & Karatsoreos, I. N. *Sleep Med. Clin.* **10**, 1–10 (2015).
17. Slats, D., Claassen, J. A. H. R., Verbeek, M. M. & Overeem, S. *Ageing Res. Rev.* **12**, 188–200 (2013).
18. Dart, A. *Nature Rev. Cancer* **16**, 409 (2016).

This article was published online on 23 November 2016.

Stem cells and interspecies chimaeras

Jun Wu^{1,2}, Henry T. Greely³, Rudolf Jaenisch⁴, Hiromitsu Nakauchi^{5,6}, Janet Rossant^{7,8} & Juan Carlos Izpisua Belmonte¹

Chimaeras are both monsters of the ancient imagination and a long-established research tool. Recent advances, particularly those dealing with the identification and generation of various kinds of stem cells, have broadened the repertoire and utility of mammalian interspecies chimaeras and carved out new paths towards understanding fundamental biology as well as potential clinical applications.

In Greek mythology, a chimaera is a fearsome fire-breathing beast composed of different parts of more than one animal, vividly depicted in Homer's *Iliad* as a lion-headed creature with the body of a goat and the tail of a serpent. In modern bioscience, chimaeras are entities made up of cells from different organisms. They are extremely valuable basic biology research tools with the potential for future clinical use. Experimental chimaeras generated from cells of more than one individual of the same species, particularly the mouse, have been widely used for many biomedical studies. Here, however, we focus on mammalian chimaeras generated from different species, also known as interspecies chimaeras (see Box 1 for related terminologies), which have recently garnered attention among researchers and the public, owing to their potential for providing replacement human organs.

This Review will first provide a historical perspective and a summary of recent advances in stem-cell-derived interspecies chimaeras and their potential applications. It will go on to discuss ethical guidelines, current policies, societal effects and, finally, conclude with an outlook on research in this and related areas. Interspecies chimaera-related research covers a broad range of topics. We focus here on studies that deliver donor stem cells to pre-natal host embryos and fetuses—research on conventional xenotransplantation approaches involving the introduction of donor cells into the post-natal animal has been reviewed elsewhere¹.

Earlier studies on interspecies chimaeras

Interspecies chimaeras in mammals were first developed in the 1970s as tools to aid the study of cell lineage and cell fate during embryonic development. Prior to the development of transgenic reporter lines, there was a need to explore other ways of identifying the origin of the different cell types brought together in a chimaera. Building on the pioneering work of Le Douarin and colleagues using chick–quail chimaeras^{2,3}, in the 1970s Gardner and Johnson were the first to test the species boundaries in mammals with rat–mouse chimaeras⁴. When rat inner cell masses (ICMs) were injected into mouse blastocysts and transferred to the mouse uterus, the detailed distribution of the rat-derived cells could be observed in sections of the fetuses using species-specific antibodies. However, when the chimaeras were left to go to term, the resulting offspring had very few detectable rat cells remaining. The reason for this loss is still not clear, but it was suggested that different developmental rates between two species, that diverged evolutionarily about 20.9 million years ago (www.timetree.org), could lead to selection against the rat cells.

Interspecies chimaeras between two more closely related murine species, *Mus musculus* and *Mus caroli*, were generated in further work

by Rossant and Frels⁵. Live chimaeras with cell contributions from both species were produced by injection of *M. caroli* ICMs into *M. musculus* blastocysts⁵. DNA–DNA *in situ* hybridization and differences in satellite DNA between the two species were used to provide a detailed description of the tissue types derived from the ICM and the trophectoderm of the blastocyst⁶. Following this success, other viable interspecies chimaeras were also generated, including sheep–goat⁷ and two bovine species, *Bos taurus*–*Bos indicus*⁸.

The successful production of interspecies chimaeras is dependent on the matching of the species origin of the trophectoderm derivatives and that of the maternal uterus. *Mus caroli* blastocysts, for example, cannot survive beyond early postimplantation in the uterus of *M. musculus*; however, viable *M. caroli* offspring can be produced from a *M. musculus* mother when a blastocyst is reconstituted with *M. musculus* trophectoderm and *M. caroli* ICM^{9,10}. This suggests that species boundaries can be extended considerably, provided that the interspecies combination is confined to the ICM derivatives.

These early studies paved the way for understanding evolutionarily conserved and divergent developmental processes in an interspecies setting *in vivo*. The derivation of pluripotent stem (PS) cells from early embryos, as well as from differentiated cells through cellular reprogramming, has renewed the interest in generating interspecies chimaeras¹¹.

Stem cells and interspecies chimaeras

The first PS-cell-derived interspecies chimaeras were generated by Lahn and colleagues by injecting embryonic stem (ES) cells of the wood mouse *Apodemus sylvaticus* into *M. musculus* blastocysts. Although *A. sylvaticus* diverged from *M. musculus* about 14.59 million years ago, viable chimaeras were obtained that contained a considerable contribution (up to 40% in some tissues) of *A. sylvaticus* cells in all major organs, including germ cells, in mice¹². Notably, *A. sylvaticus*–*M. musculus* chimaeras displayed features that were intermediate between the two species, including eye size and jumpiness, although these phenotypes were not statistically quantified¹². Whether or not interspecies chimaeras can retain behavioural traits characteristic of the donor species is an intriguing question. A recent study by Goldman and colleagues has suggested that the extensive contribution of human glial progenitors after grafting into the post-natal mouse forebrain could alter the cognitive capability of the host¹³.

Improvement in mouse ES cell culture eventually enabled the derivation of germline-competent rat ES cells^{14–16}, rekindling the interest in generating rat–mouse chimaeras. A milestone study by Nakauchi and colleagues reported the generation of viable adult rat–mouse chimaeras

¹Salk Institute for Biological Studies, 10010 North Torrey Pines Rd, La Jolla, California 92037, USA. ²Universidad Católica San Antonio de Murcia (UCAM) Campus de los Jerónimos, nº 135 Guadalupe 30107, Murcia, Spain. ³Center for Law and the Biosciences, Stanford Law School, 559 Nathan Abbott Way Stanford, California 94305-8610, USA. ⁴The Whitehead Institute for Biomedical Research and Department of Biology, MIT, Cambridge, Massachusetts 02142, USA. ⁵Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, California 94305-5101, USA. ⁶Center for Stem Cell Biology and Regenerative Medicine, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. ⁷Program in Developmental and Stem Cell Biology, Hospital for Sick Children Research Institute, 686 Bay Street, Toronto, Ontario M5G 0A4, Canada. ⁸Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada.

BOX 1

Defining chimaeras

A chimaera is typically defined as an organism composed of a mixture of different cell populations that derive from more than one zygote. They can be formed by processes such as the mixing of early embryos or the grafting of tissues from different stages of development or the adult. Chimaeras are distinct from hybrids (offspring resulting from breeding between different species) and from mosaics (animals carrying genetically different cells that originated from the same zygote). Chimaeras can be categorized as intraspecies or interspecies, depending on whether cell derivatives from two zygotes are from the same or different species, respectively.

Creating interspecies chimaeras requires a donor species and a host species. The donor species provides cells of embryonic, fetal, or adult origin, either from primary tissue or from cell cultures. The host species provides the physiological environment and life support for embryonic, fetal, or adult chimaeric animals. In stem-cell research, the generation of an interspecies chimaera often involves the transplantation of multi- or pluripotent stem (PS) cells from the donor into an animal recipient at embryonic, fetal or post-natal stages of development.

Tissue distribution and duration of chimaerism often differ depending on the donor cell type and the developmental stage of the host. PS cells are less restricted in developmental potential than other stem-cell types and thus can give rise to a high degree of chimaerism with wide tissue distribution when transferred to pre-implantation host embryos. Chimaeras generated by progenitors or adult stem cells are more confined to specific tissues and may need to be grafted to appropriate sites and developmental stages. Chimaeras are often designated according to the developmental age of the host at the time of analysis (for example, chimaeric embryos and chimaeric fetuses) or at the time of donor cell injection (for example, blastocyst chimaeras).

Other important distinctions are between heterotopic and orthotopic, and heterochronic and isochronic chimaeras. Orthotopic chimaeras are generated by transplanting donor cells to their cognate location where they can participate in natural developmental processes or proper tissue organization of the host (for example, transplanting donor hepatocytes to the host liver). Heterotopic chimaeras occur with differentiation or integration of donor cells at a different site within the host animal than their origin (for example, transplanting donor pancreatic β -cells to the host liver). Heterochronic and isochronic chimaeras are distinguished by the temporal properties between donor and host. If donor cells are delivered to the host at a time matching their *in vivo* origin, an isochronic chimaera will be formed, otherwise the chimaera formed will be heterochronic.

by the injection of rat PS cells into mouse blastocysts or by injection of mouse PS cells into rat blastocysts¹⁷. Chimaerism in interspecies fetuses varied among individuals and tissues, but was lower than that in the mouse–mouse or rat–rat intraspecies chimaeras. The production of viable rat PS-cell-derived rat–mouse chimaeras has been confirmed by other studies (ref. 18 and J.W. *et al.*, manuscript in preparation). Wu *et al.* further showed that rat PS-cell-derived rat–mouse chimaera aged normally and lived an entire mouse life span (about 2 years) (J.W. *et al.*, manuscript in preparation). The higher chimaerism rate of rat PS cells versus rat ICMs suggest that rat PS cells may have growth advantages over rat ICMs, perhaps as a result of their *in vitro* culture. The phenotype of rat–mouse chimaeras provides several insights. First, the chimaeras born from rat surrogates were generally rat-sized and those born from mouse surrogates were mouse-sized unless there were high contributions by the xenogenic PS cells, a factor which was typically associated with morphological abnormalities and embryonic lethality¹⁷. Second, the chimaeric contribution of xenogenic cells seems to display lineage bias. Although a high percentage of mouse CD45⁺ cells were detected in rat-blastocyst-derived chimaeric livers, rat CD45⁺ blood cells were rarely detected in mouse-blastocyst-derived chimaeric fetal livers¹⁷. This observation was specific to interspecies chimaeras, as the same rat PS cell line efficiently contributed to CD45⁺ cells in intraspecies chimaeric fetal livers¹⁷. Third, as mice have a gallbladder whereas rats do not, the presence or absence of the gallbladder in rat–mouse chimaeras appeared to be determined by the host species, suggesting that the donor PS cell derivatives are subject to regulation by the host programs that

drive organogenesis¹⁷. Intriguingly, Wu *et al.* recently found rat PS cells could contribute to mouse gallbladder, suggesting that the mouse embryonic microenvironment was able to unlock a gallbladder developmental program in rat cells that is normally suppressed during rat development (J.W. *et al.*, manuscript in preparation). Whether or not high contributions of rat PS cells to mouse-blastocyst-derived chimaeras would interfere with mouse gallbladder development remains an interesting, unresolved issue.

Unlike rodents, germline-competent ES cells have yet to be isolated from other species. Stable ES cell lines could be derived from both non-human primate and human blastocysts^{19–24}; however, not one of these lines was able to generate intra- or interspecies blastocyst chimaeras^{25–27}. It was later realized that the rodent and primate ES cells represent different phases of pluripotency, naive and primed, respectively²⁸. Mouse ES cells represent the naive epiblast state²⁹, whereas epiblast stem cells (EpiS cells), another PS cell type derived from the post-implantation epiblast of rodent embryos^{30,31}, are primed for differentiation. Many defining features of EpiS cells were frequently found in primate ES cells, suggesting that primate ES cells were stabilized at the primed pluripotent state. Consistent with their post-implantation identity, mouse EpiS cells can engraft into egg cylinders in culture and differentiate into all three embryonic germ lineages, including primordial germ cells^{32–34}. Similarly, human PS cells grown in different primed cultures can also integrate into mouse gastrula-stage embryos and form *ex vivo* interspecies chimaeras^{34,35}. Izpisua Belmonte and colleagues have since shown that primed human PS cells could chimaerize gastrula stage chick embryos as well (J.W. *et al.*, unpublished data). Thus, gastrula-stage embryos, rather than pre-implantation blastocysts, permit primed human PS cells to engraft and cross the xenobARRIER^{36,37}.

With the recognition that primate ES cells correspond to the primed state, an unresolved question was whether human PS cells that were more similar to mouse ES cells could be obtained. Derivation of genuine naive human ES cells, if achieved, would provide several practical advantages over primed cells, including improved single-cell cloning efficiency, ease of genome editing and the potential to generate interspecies chimaeras. The first successful attempt at generating mouse-ES-cell-like human ES cells came from a study by Jaenisch and colleagues in which the forced expression of *Oct4* (also known as *Pou5f1*), *Klf4* and *Klf2* transcription factors was used to revert primed human ES cells to a more immature state. Although converted cells share many salient features of mouse ES cells, their long-term maintenance was still dependent on ectopic transgene expression³⁸. Following this pioneering work, a flurry of recent studies have reported conditions that produce stable, transgene-free immature human PS cells with molecular features resembling mouse ES cells^{39–47}. Some of these cultures have also been used for the *de novo* derivation of human ES cells from human blastocysts^{41,43,46,47}.

It has proven difficult to define the gold standard criteria for assigning the true naive human pluripotent state⁴⁸. Fan and colleagues used a systems biology approach to assess the gene networks of different PS cells and found that the gene networks of mouse ES cells and EpiS cells were reproducible across all datasets examined; however, naive human PS cells grown in different cultures exhibited high degrees of variation⁴⁹. Naive gene networks appeared to be poorly conserved between human and mouse PS cells, and both better resembled their respective blastocysts⁴⁹. Thus, a direct comparison of naive human PS cells to pre-implantation human embryos can potentially serve as a more reliable molecular criterion for defining the human naive pluripotent state. In this regard, a recent study showed that the gene expression profiles of naive 5i- and T2iL-Gö-conditioned human PS cells better correspond to that of the human cleavage-stage embryo than cells produced by other protocols⁴⁴.

Validation of the naive state pluripotency in the mouse is achieved by demonstrating the functional contributions made by the mouse PS cells to blastocyst chimaeras. Owing to ethical considerations, the testing of human putative naive cells in such assays is limited to the use of animal host embryos, usually the mouse. Initial attempts at generating human–mouse chimaeric embryos using naive human PS cells highlight the potential difficulties that lie ahead in establishing interspecies chimaeric

models. Hanna and colleagues reported the robust generation of embryonic day (E)10.5 human–mouse chimaeric embryos using human PS cells grown in NHSM medium⁴¹. By contrast, there were several other studies that contradicted this conclusion, outlining the inefficient, limited chimaeric contribution of cells cultured in NHSM medium or other naive conditions^{42–44,50}. Notably, though, non-human primate PS cells cultured in modified NHSM cultures were reported to contribute to the formation of pre-natal monkey–monkey and monkey–mouse chimaeras, albeit to a modest degree^{51,52}.

In most studies, the criteria used to evaluate chimaerism were exclusively based on the presence of fluorescently labelled human cells. There are several potential problems associated with this, especially when the degree of chimaerism is low, including auto-fluorescence and the uptake of markers from dying cells. A recent study, by Theunissen *et al.*⁴⁴ used both fluorescence detection and quantitative PCR analysis of human mitochondrial DNA (sensitive enough to detect a single human cell in 10,000 mouse cells), to analyse 1,400 putative human–mouse chimaeric embryos, finding only 7 to be truly chimaeric⁴⁴. Those that were chimaeras contained between 1 and 4 human cells per 10,000 mouse cells. Although the absence of human mitochondrial DNA is a strong negative result, the presence of a positive signal does not necessarily indicate the functional integration of donor cells as DNA can be derived from dead cells or contamination during sample preparations. Similarly, the presence of GFP-positive donor cells alone is not definitive proof for chimaeric contribution as human PS-cell-derived cells often appear as a cluster and do not fully integrate into animal tissues. Definitive proof of human–animal chimaera formation therefore requires evidence of functional contributions.

Humans and mice differ considerably in various aspects, including post-implantation epiblast development, embryo size, speed of development and gestational period. These and other differences may affect the integration, proliferation and differentiation of human PS cells in the chimaeric epiblast. Interspecies chimaera research with large animal hosts that are more similar to humans in anatomy, physiology and organ size could result in an improved research model. Experiments evaluating the chimaeric contribution of various human PS cells in large animal hosts are currently lacking. Izpisua Belmonte and colleagues have injected several existing human PS cell lines into pig blastocysts followed by embryo transfer to surrogate sows and results seem to agree with those from human–mouse chimaeric studies (J.W. *et al.*, manuscript in preparation). Interestingly, both rat and mouse PS cells, which can robustly generate interspecies chimaeras among rodents, failed in chimaeric contribution to day E21–E28 pig embryos (J.W. *et al.*, manuscript in preparation). Parallel attempts by Nakauchi and colleagues to inject human PS cells into sheep morulae/blastocysts have found a similarly limited human chimaeric contribution 28 days after transfer into surrogate ewes (T. Rashid *et al.*, unpublished data).

In addition to PS cells, more developmentally restricted stem/progenitor cells are also amenable to chimaera formation. Neural crest cells (NCCs), a multipotent embryonic progenitor cell population, first emerge in the neural plate border region, between the neural epithelium and epidermis during neurulation⁵³. Jaenisch and colleagues injected NCCs derived from E8.5 mouse embryos, as well as those from mouse ES cells, into the amniotic cavity of E8.5 mouse embryos *in utero*^{54,55}. This allowed the NCCs to enter the neural tube through the open neuropore, thereby contributing to chimaeric pigmentation formation in post-natal mice^{54,55}. More recently, the authors also generated interspecies neural crest chimaeras after transplanting PS-cell-derived rat and human NCCs into E8.5 mouse embryos *in utero*. Notably, both the rat–mouse and human–mouse neural crest chimaeras were born and the donor cells differentiated and matured into functional melanocytes and contributed to visible hair pigmentation in the adult mouse⁵⁶.

Human neural stem cells have also been transplanted into rat fetuses to generate interspecies chimaeric brains^{57,58}. Transplanted human neural stem cells displayed wide chimaeric contribution to the rat central nervous system and gave rise to neurons, astrocytes and oligodendrocytes. Human haematopoietic stem cells (HSCs) have been routinely transplanted

into immunodeficient mice to assess their function and differentiation capacity⁵⁹; however, this is mostly by post-natal injection. Human–sheep and human–pig haematopoietic chimaeras, generated via the *in utero* transplantation of human CD34⁺ haematopoietic stem progenitor cells, have been reported but the contribution of human blood cells was limited in these experiments—probably because host HSCs competed with transplanted cells for the stem-cell niche^{60,61}. A summary of chimaeras generated using different types of cultured stem cells is provided in Fig. 1.

Factors involved in interspecies chimaerism

To improve the degree of chimaerism achieved from the transplantation of xenogeneic stem cells, especially PS cells, several important factors need to be considered (Fig. 2).

Matched developmental timing

Transplantation of donor cells into developmental-stage-matched (isochronic) host tissue may be critical for the efficient engraftment of cells into chimaeras. Mouse ES cells were able to gain entry into host development following blastocyst injection but failed to thrive in the post-implantation epiblast³². Conversely, although mouse EpiS cells are inefficient at generating blastocyst chimaeras, they are capable of colonizing the E7.5, but not the E8.5, epiblast, a stage at which pluripotency has been lost³². Similar observations have been made for human PS cells. Naive human PS cells were readily incorporated into mouse^{41,42,50}, pig and cow (J.W. *et al.*, manuscript in preparation) ICMs but later contributions to the developing fetuses were rare and inefficient (refs 42–44, 50 and J.W. *et al.*, manuscript in preparation). Conversely, primed human PS cells efficiently engrafted into E6.5 or E7.5 mouse epiblasts *in vitro*, but not pre-implantation mouse^{27,34,35}, pig and cow (J.W. *et al.*, manuscript in preparation) blastocysts. Several studies reported heterochronic chimaeras, in which cells of early developmental ages were able to respond to a later tissue environment. Gage and colleagues⁶² injected human ES cells into the lateral ventricle of E14 mice. The E14 mouse brain allowed some injected human ES cells to differentiate and generate mature, active human neurons in the adult mouse forebrain, unlike the post-natal brain, in which human ES cells generate teratomas⁶². Shinohara and colleagues generated functional spermatozoa after transplantation of E6.5 epiblast or E8.5–E16.5 mouse fetal germ cells into the seminiferous tubules of a post-natal mouse testis⁶³. The reverse experiments, in which more advanced cells were returned to earlier-stage embryos, were less successful. Primordial germ cells isolated from post-implantation E8.5–E11.5 embryos did not contribute to chimaera formation following blastocyst injection⁶⁴. Several groups have reported the transplantation of various differentiated lineages—such as neural precursor cells, HSCs or mesenchymal stem cells—into blastocysts and have claimed that the injected cells contributed to the respective lineages^{65–68}. The validity of these findings has, however, not been confirmed and remains controversial.

Providing selective advantage for donor cells

Several strategies may confer a selective growth advantage to donor cells. First, an empty host niche strategy. If the host is genetically modified, compromising or eliminating the development of certain cell lineage(s) in embryogenesis, the recipient niche can be used exclusively by donor cells for their differentiation, proliferation and function. This strategy has been widely used for assaying HSC functions *in vivo* after the myeloablative irradiation of a donor's bone-marrow stem-cell niche⁶⁹. Spermatogonial stem cells or primordial germ cells can be injected into the seminiferous tubules of a recipient *W/W^V* mouse testis lacking endogenous germ cells for proper spermatogenesis^{63,70,71}. *c-Kit*-mutant *W^{sh}/W^{sh}* mice that lack melanoblasts are more permissive for the chimaeric contribution of NCCs derived from rodent and human PS cells⁵⁶. Notably, this concept has also been adapted for the generation of xenogenic organs via interspecies blastocyst complementation (refs 17, 18 and J.W. *et al.*, manuscript in preparation).

Second, a cell-competition strategy. Cell competition was first studied in *Drosophila* in which cells carrying a *Minute* (also known as

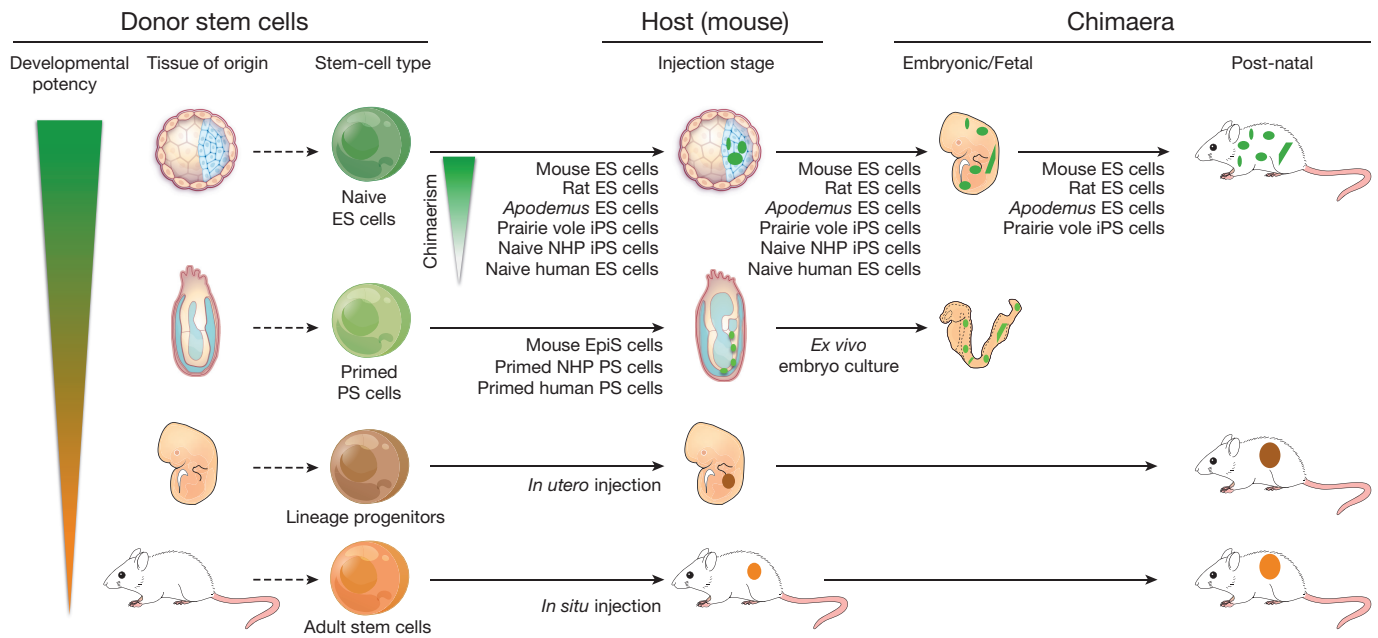


Figure 1 | Chimaeras generated from cultured stem cells. Stem cells with varying degrees of developmental potency have been derived from different developmental stages. *In situ* transplantation allows cultured stem cells to generate intra- or interspecies chimaeras. Rodent ES or iPS cells can contribute to post-natal chimaeras in a host from the same or a different rodent species. Primate naive PS cells can contribute, albeit

to a very limited degree, to developing mouse embryos after blastocyst injection. Primed PS cells can engraft into mouse gastrula stage and generate *ex vivo* chimaeric embryos. Lineage progenitors or adult stem cells can contribute to chimaera formation in an organ of a developing fetus or post-natal animal. NHP, non-human primate.

RpS17) mutation were outcompeted by wild-type cells with metabolic advantages⁷². Later studies in mammalian systems revealed that this process is universal and highly conserved⁷³. MYC-induced super-competition constitutes another mode of cell competition, in which cells with higher MYC expression out-compete neighbouring wild-type cells⁷³. Both types of cell competition have, thus far, only been examined in the intraspecies setting and their roles in interspecies chimaera formation await exploration. Overexpression of c-MYC in human-PS-cell-derived NCCs did not seem to give a more competitive edge for human cells, suggesting that super-competition conferred by c-MYC may not work across species⁵⁶. Zwaka and colleagues identified a network of genes whose downregulation confers embryonic cells with the ability to out-compete wild-type cells in development, a feature reminiscent of MYC-driven super-competition⁷⁴. Whether or not expression of these pro-competition genes would promote cross-species contributions remains to be determined.

Third, an enhancement of donor cell survival approach. Nakauchi and colleagues found cells from a subclone of EpiS cells could generate blastocyst chimaeras, in part owing to their resistance to cell death after injection into blastocysts⁵⁰. This led to further examination of whether the forced expression of the anti-apoptotic protein BCL2 might allow primed PS cells to engraft in blastocysts, survive and contribute to chimaeras. Indeed, BCL2-overexpressing EpiS cells also survived in mouse blastocysts and contributed to chimaeric mice⁷⁵. Furthermore, when BCL2-expressing SOX17⁺ endoderm progenitors were injected into blastocysts, they also contributed to chimaera formation, but only to gut tissues⁷⁵. These results suggest that the prevention of apoptosis supports the survival of grafted progenitors in pre-implantation embryos and can extend the degree of heterochrony that can be tolerated in blastocyst chimaeras. Once the appropriate differentiation stage is reached, surviving progenitors will then take part in embryogenesis and follow their ordained developmental fate. The prevention of apoptosis is also a valid strategy for mouse–rat interspecies chimaera formation. For example, BCL2-expressing rat primed PS cells can contribute to chimaeras after injection into mouse blastocysts⁷⁵. BCL2 overexpression also promoted human ES cell survival in mouse embryos

in vitro (H. Masaki *et al.*, unpublished data). Similar results were obtained with BCL2-overexpressing monkey ES cell lines in mouse embryos *in vivo* (H. Masaki *et al.*, unpublished data). Although the data are promising, it should be noted that the progeny of primate PS-cell-derived cells were diverted towards the extraembryonic lineage. Thus, more experiments and optimizations are warranted before any definitive conclusions can be drawn.

Evolutionary considerations

Overall, the generation of interspecies chimaeras between mouse and rat is less efficient than the generation of intraspecies chimaeras. The cause of this ‘xenobarrier’ is not clear; however, there are several potential reasons, including ligand–receptor incompatibilities and differences in the affinity of adhesion molecules as a result of genetic diversification. To gain evolutionary insights into the xenobarrier, Nakauchi and colleagues generated induced pluripotent stem (iPS) cells from the prairie vole (*Microtus ochrogaster*), a rodent species that diverged from *M. musculus* about 32.7 million years ago. When prairie vole iPS cells were injected into mouse embryos, interspecies chimaeras could be generated, but the embryonic development rate, degree of chimaerism and survival to adulthood were lower than in mouse–rat chimaeras (H. Sato *et al.*, manuscript in preparation). These data imply that genetic diversification or evolutionary distance is at least partly responsible for the xenobarrier. In line with this, Jaenisch and colleagues observed a greater chimaeric contribution of mouse NCCs to the mouse embryo than that made by rat NCCs, and rat NCCs showed more efficient engraftment than human NCCs⁵⁶.

Potential applications

Interspecies chimaeras are excellent experimental models for studying development, organismal homeostasis, stem-cell potential and disease. Recently, progress in several technological frontiers has opened new possibilities for interspecies chimaera research. Through genetic manipulation, the developmental niche(s) of the host animal can be made to serve donor cells exclusively, thereby generating organ- or tissue-enriched chimaeras, which at some stage could help provide much-needed organs

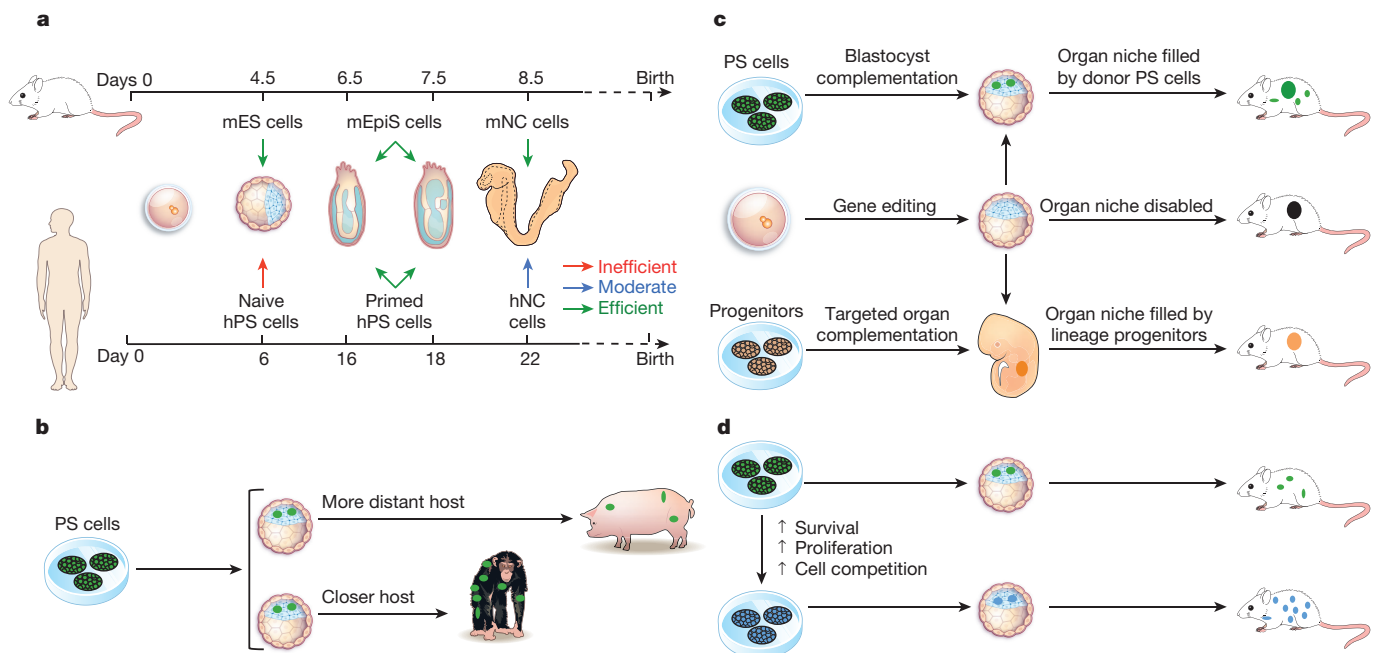


Figure 2 | Strategies to improve interspecies chimaerism with human stem cells. **a**, Matching developmental timing is critical for pluripotent and multipotent stem cells to engraft and generate intra- and interspecies chimaeras. **b**, Choosing a host that is evolutionarily closer may help increase the degree of chimaerism by donor human PS cells. **c**, The genetic

modification of host animals to disable organ development may enable donor human PS cells or progenitors to populate the targeted organ with minimal competition from the host. **d**, The modification of donor human PS cells to enhance their survival, proliferation and/or cell competition may also help increase the degree of chimaerism.

for transplant. Interspecies chimaeras with human stem-cell contributions could also serve as a platform for disease modelling and drug screening, providing *in vivo* readouts of disease onset and progression, drug efficacy and toxicity, with relevant clinical value¹¹ (Fig. 3).

Organ generation

This can be potentially approached with blastocyst complementation or targeted organ complementation.

Interspecies blastocyst complementation. The ability to grow entire body parts for patients in the laboratory, as replacements for damaged or failing organs, is the holy grail of regenerative medicine. This quest took a considerable step forwards ten years ago with the discovery of iPS cells⁷⁶. Although efforts have been made to generate functional mature cells from iPS cells that target diseases that are potentially amenable to treatment by cell transplantation, other conditions, such as heart, kidney, liver and lung failure, require whole-organ replacement. Generating such complex, three-dimensional tissues from iPS cells once seemed impossible. In 2010, Nakauchi and colleagues reported the generation of a rat pancreas in mouse, using rat iPS cells to complement the pancreatic organ niche in mouse embryos that are unable to develop a pancreas¹⁷. More recently, the same group succeeded in generating a mouse pancreas in *Pdx1*^{-/-} rats (T. Yamaguchi *et al.*, manuscript in preparation). To evaluate functionality, mouse islets formed in rats were transplanted into mice with streptozotocin-induced diabetes. Although the isolated islets contained some rat endothelial cells and other non-parenchymal cells, 100 islets transplanted under the mouse renal capsule normalized blood glucose levels over 370 days without immunosuppression (T. Yamaguchi *et al.*, manuscript in preparation). This proof-of-concept study indicates that organs generated from iPS cells in this manner can be just as functional as those from wild-type animals. In addition to pancreas, Isotani *et al.*¹⁸ complemented blastocysts from nude mice lacking a thymus with rat ES cells and generated an xenogenic rat thymus. Usui *et al.* tested the blastocyst complementation strategy for the generation of kidney tissue in *Sall1*-knockout mice. When mouse PS cells were used, a kidney was successfully generated. However, rat PS cells failed in this context, suggesting that key molecules involved in the interaction between mesenchyme and ureteric

bud during kidney development might not be conserved between mice and rats⁷⁷. Most recently, Izpisua Belmonte and colleagues broadened the utility of interspecies blastocyst complementation by combining CRISPR–Cas9-mediated zygote genome editing with the blastocyst injection of rat PS cells, thereby removing the need for existing mutant mouse strains (J.W. *et al.*, manuscript in preparation). This strategy allowed successful complementation for several lineages, including pancreas, heart and the eye, among others (J.W. *et al.*, manuscript in preparation).

These proof-of-concept studies could perhaps be expanded beyond rodents to larger animals, whose organs are much more similar in size to those of humans. That is, pig, sheep or monkey embryos may eventually be complemented with human stem cells to generate replacement organs for any part of the human body. Pig–pig blastocyst complementation has already been achieved by Nakauchi and colleagues⁷⁸. Pig fibroblasts overexpressing the transcriptional repressor HES1 under the pancreas-specific *PDX1* promoter were cloned to give rise to embryos carrying a *PDX1-HES1* transgene⁷⁸. *PDX1-HES1* expression suppresses the pancreatic development program, thus leading to the creation of pancreatogenesis-disabled pig embryos. Owing to the lack of chimaera-competent pig PS cells, cloned blastomeres expressing the fluorescent protein huKO were used to complement the *PDX1-HES1* embryos. These huKO⁺ blastomeres were able to contribute to chimaera formation, generating an entire huKO⁺ pancreatic epithelium. Moreover, the chimaeric pigs generated by complementation were able to grow into adulthood with a functional pancreas. Whether or not human PS cells can generate xenogeneic organs in pigs remains unknown. Although similar in physiology and organ size, the evolutionary distance between human and pig (96 million years) is even greater than that between human and mouse (90 million years). Thus, the chimaeric contribution made by human PS cells is expected to be low in pigs, consistent with recent findings (J.W. *et al.*, manuscript in preparation). Choosing a species closer to humans, such as non-human primates, may be more likely to work, although the law in many countries prohibits such experiments. Developing strategies to improve human chimaerism in a distant animal host as discussed above, could help to turn this goal into reality in the future.

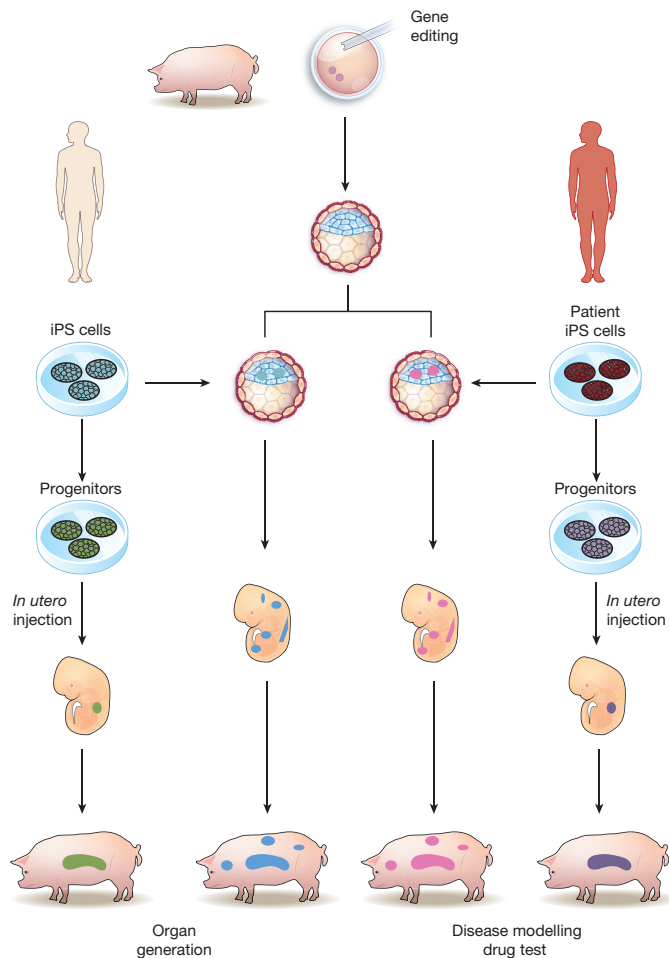


Figure 3 | Potential application of interspecies chimaeras with human stem cells. Organ generation via interspecies blastocyst complementation or interspecies targeted organ complementation could help to solve the severe shortage of organ donors worldwide. Human–animal chimaeras will also help researchers to understand the aetiology, onset and progression of human diseases better. Additionally, they could be useful for testing the efficacy and safety of a candidate drug *in vivo*.

Interspecies targeted organ complementation. One of the issues regarding *in vivo* organ generation in animals is ethical concern over potential gamete and neural contribution from PS cells. This issue could be addressed by developing methods of targeted organ generation, minimizing the likelihood of any undesired gameto- or neurogenesis⁷⁹. One approach is to modify PS cells, restricting their differentiation to the organ of interest. Conditional expression of the transcription factor gene *Mixl1* made it possible to induce differentiation of PS-cell-derived cells towards the endodermal lineage, thereby reducing the number of non-endodermal PS-cell-derived cells⁸⁰. The introduction of constructs containing suicide genes under neural- or germ-cell-specific promoters to eliminate completely the formation of human iPS-cell-derived neural and germ cells in the host animal could be explored. Another approach is the use of committed progenitors or organ buds instead of PS cells. As discussed, the organ niche of organ-deficient embryos is available for complementation throughout development of the organ. Therefore, injecting lineage-committed progenitor cells, (such as cultured nephron progenitor cells⁸¹, rather than PS cells, into the embryo at the right place and time may allow complementation of organ deficiency with little chance of generating ‘off-target’ humanized tissues.

If the host organ niche is emptied by genetic manipulations, donor human progenitor cells will probably be able to colonize it, giving rise to a human-cell-enriched organ. Although theoretically this is a sound strategy and potentially reduces the ethical concern, there are caveats: the

developmental stages need to be matched between human progenitors and the host animal fetus; delivering human cells before the host organ becomes atretic is likely to be critical and challenging; and delivery of human progenitor cells needs to be performed before the host immune system is formed (less than 50 days in pigs).

In vivo disease modelling

An important goal of experimental biology is the development of model systems that allow for the study of human diseases under *in vivo* conditions. Transgenic animals have been successful in modelling a variety of human diseases but have failed to provide disease-appropriate phenotypes for monogenic diseases such as Lesch–Nyhan syndrome⁸² or of complex diseases such as Parkinson’s or Alzheimer’s. Given that disease-specific iPS cells carry all genetic alterations that contributed to disease pathology, the functional integration of patient-derived cells into the tissues of the developing mouse embryo would allow for study of the initiation, progression and manifestation of the disease in question. Host-specific developmental and physiological programs may alter the behaviour of donor human cells in a non-human host. Nevertheless, the ability to study the autonomous versus non-autonomous effects of the multiple genetic alterations that contribute to a human disease under *in vivo* conditions cannot be dismissed.

In vivo drug screening

Compared to existing drug screening platforms, which often involve the use of patient samples and immortalized cell lines and/or *in vivo* transgenic mouse models, the use of interspecies chimaeras with human stem cells could potentially offer a superior *in vivo* drug-screening platform. Interspecies chimaeric formation using human iPS cells or iPS-cell-derived progenitors offers an attractive platform for the personalized *in vivo* testing for drug efficacy and toxicity. This approach holds the potential to be a robust preclinical testing platform for more accurate predictions of clinical outcomes.

Ethical, legal and social issues

In his January 2006 State of the Union address, President George W. Bush asked Congress to pass the ‘Human Chimaera Prohibition Act’ that would have made many activities that involve the mixing of human genes, cells, or tissues with those of non-human animals illegal⁸³. Ten years later, political, ethical, and social concerns about chimaeras remain.

The issues

Four major ethical and social issues exist concerning research with, and potential human clinical use of, chimaeras: animal welfare, sources of donor cells, general public discomfort with chimaeras, and the ‘humanization’ of the host species.

With chimaeras, as with any new intervention with non-human animals, the pain or disability the animals could suffer may be hard to predict. Researchers must therefore be more sensitive to the possibility of pain and suffering than with existing, well-characterized laboratory animals^{84,85}. If chimaeras were ultimately developed to provide transplants for humans, carnivorous societies should not ban their use, but their humane treatment and slaughter would be essential.

Although widely legal for research use, some people consider the use of human ES cells or of cells derived from aborted fetuses unethical; this would be true for chimaeras as well.

Public discomfort, sometimes expressed as concerns about ‘unnaturalness’, clearly exists around chimaeras. These seem to boil down to a belief that any species mixing is inherently wrong⁸⁶. Whether public concerns without a good logical basis⁸⁷ should be considered ‘ethical’ issues is debatable, but this could have a considerable effect on research and its funding.

The most controversial aspect of human–non-human chimaeras has been the fear that they may in some way confer ethically important human characteristics on non-human hosts. Substantial discussion of the humanization concern started around 2003 (ref. 88) and has continued for several years^{84,85,89–91}. The precise concerns about humanization have

remained unclear. One author argues that three specific characteristics of human–non-human chimaeras should be carefully examined: human-like cognitive abilities, human gametes, or human-like appearances⁹². Most discussion has focused on the brain; some suggest that chimaeras could lead to moral confusion of humans and non-humans⁸⁸ or diminution of human dignity^{89,90,93,94}. One powerful, albeit unlikely, concern is that such chimaeras could achieve sufficient human-like consciousness or intelligence to ‘deserve’, but be denied, personhood⁹¹. Others merely note that the issue is controversial and argue that researchers should be very careful with research that could be seen as possibly giving the host some human-like cognitive abilities^{84,85}.

As parts of this Review have indicated, the creation of a non-human animal with a substantially human brain, or with enough human characteristics to warrant human treatment⁹⁵ seems, at present, unlikely—especially given the possible measures to limit unplanned spread of the stem cells. The possibility cannot be entirely dismissed; one study has already shown improvements in mouse memory after insertion of human glial progenitor cells¹³ and the increasing use of naive human PS cells may lead to wider differentiation.

Laws, regulations and guidance

Different national laws and regulations concerning laboratory animal care and use exist, as they do for proper treatment of animals reared for food. These need only to be carefully applied to novel chimaeric animals. Similarly, varying national laws and regulations govern the use of fetal tissue or human ES cells in research. Because chimaera research might produce dramatic results, public attention could easily be brought to these issues. The following case, concerning the United States, may indicate how some of these concerns may be handled.

In the late 1980s, the United States banned the federal funding of research with fetal tissue⁹⁵. The 1993 NIH Revitalization Act⁹⁶ authorized federal funding for research involving appropriately obtained fetal tissue. Recent controversy over Planned Parenthood’s acquisition of fetal tissue for transfer to researchers has revived controversy about this statute.

Similarly, federal funding for research that could create human ES cells has been banned since 1995. In 2000, the Clinton administration concluded that this allowed federal funding of research to study human ES cell lines that others had created using non-federal funding. The Bush administration stringently limited the extent of such funding, although the Obama administration broadly permitted it in 2009—a position that was upheld by a federal appellate court⁹⁷. The position, however, can be changed by legislation or by a policy decision introduced by a new president. (Note that, in the United States, this issue has mainly concerned federal funding for research, not limitation of non-federally funded research.)

Of the ethical concerns surrounding the generation of chimaeras, humanization has received the most attention, largely through guidance rather than law or regulation. Human–non-human chimaera research has been addressed as part of general guidelines from the following bodies: the US National Academy of Sciences (NAS) in 2005 (ref. 98); the International Society for Stem Cell Research (ISSCR), first in 2006 (ref. 99), then updated in 2016 (ref. 100); and the UK Academy of Medical Sciences in 2011 (ref. 101). Specific guidance on chimaera research came from the ISSCR ethics committee in 2007 (ref. 85), while a group based at Johns Hopkins provided ethical recommendations for research inserting human cells into the brains of non-human primates¹⁰². The guidelines urge that the insertion of human stem cells (or tissues) into non-human animals be reviewed by oversight committees. Under the NAS guidelines, when a committee reviews a protocol, “particular attention should be paid to the probable pattern and effects of differentiation and integration of the human cells into the non-human animal tissues”. The NAS guidelines also ensure that the animals that have received human cells or tissues are not allowed to breed, for fear that they may produce human gametes¹⁰³. These guidelines also contain some useful specific prohibitions, such as those on placing any ES cells in human blastocysts or human ES cells into the blastocysts of non-human primates.

If properly applied, these guidelines should help to assuage concerns regarding humanization. After considering the likely patterns of differentiation and distribution, the overseeing committees should not approve research that has a considerable chance of potentially conferring human cognitive characteristics onto chimaeras or creating a human appearance.

No regulations or guidelines directly address unspecified public concerns about chimaeras. One group urged that researchers react to these worries by being open and transparent about their research, while regularly making the case that such research is important⁸⁴.

The real consequences of ethical and social concerns about chimaeras are not ancient history. In September 2015 the US NIH announced a moratorium on funding any research in which human pluripotent cells or human neural progenitor cells were placed into any non-human vertebrate embryo before gastrulation¹⁰⁴; however, the reasons for this were unclear. In November 2015, the NIH held a workshop to discuss these issues and, in August 2016, issued a draft policy calling for an internal NIH committee to review applications for some types of human–non-human chimaera funding^{105,106}. The recent US presidential election adds considerably uncertainty to the future of this line of research.

Conclusion and outlook

Notwithstanding any ground-breaking advances in generating animal interspecies chimaeras, there are currently more questions than answers concerning human–non-human chimaeras. It is not known whether more extensive chimaerism can be obtained between humans and other more related species, nor is it known whether human organs generated in chimaeras, with host nerves and vasculature, will be suitable for transplantation. In this regard, strategies such as the humanization of host animals and/or multi-lineage complementation may help. Fundamental problems remain, such as how best to resolve heterochronic developmental processes and inherent differences in gestation length between humans and other species; whether or not xeno-generated human cells, tissues and organs would be functionally compatible with human physiology; and whether it is possible to generate human cells that are able to compete equally with host cells.

The development of strategies to enhance the degree of human–animal chimaerism, such as those described in this Review, will be necessary if interspecies chimaeras are to reach their full biological and clinical potential. Engineering approaches to confer human stem cells with novel genetic circuits that monitor and control specific cell behaviours may help to overcome some of the present human–animal interspecies barriers. Similarly, engineering approaches to humanize animal models may help to lower the threshold of xenotransplantation. If successful, these new approaches, rooted in the fields of cell and developmental biology, may expand the breadth of chimaera research from the laboratory into potential clinical applications. This could aid the development of new drug screening, efficacy and toxicity methodologies, as well as the creation of disease models that may ultimately enhance diagnosis and improve the treatment of numerous pathologies.

Although the way ahead has many scientific, medical, ethical, political, financial, and other challenges and not everything that can be done in the field of chimaera research should be done, we owe it to future generations of patients and scientists to think about these challenges and experimentally proceed forward with consensual ethical, legal and social guidelines.

Received 17 June; accepted 31 October 2016.

1. Ito, R., Takahashi, T., Katano, I. & Ito, M. Current advances in humanized mouse models. *Cell. Mol. Immunol.* **9**, 208–214 (2012).
 2. Le Douarin, N. M. The ontogeny of the neural crest in avian embryo chimaeras. *Nature* **286**, 663–669 (1980).
 3. Douarin, N. L. & Kalcheim, C. *The Neural Crest* (Cambridge Univ. Press, 1982).
 4. Gardner, R. L. & Johnson, M. H. Investigation of early mammalian development using interspecific chimaeras between rat and mouse. *Nature* **246**, 86–89 (1973).
 5. Rossant, J. & Frels, W. I. Interspecific chimeras in mammals: successful production of live chimeras between *Mus musculus* and *Mus caroli*. *Science* **208**, 419–421 (1980).
- A pioneering study reporting the first viable interspecies chimaeras generated by injection of *M. caroli* ICMs into *M. musculus* blastocysts.**

6. Rossant, J., Viji, M., Siracusa, L. D. & Chapman, V. M. Identification of embryonic cell lineages in histological sections of *M. musculus* in-equilibrium *M. caroli* chimaeras. *J. Embryol. Exp. Morphol.* **73**, 179–191 (1983).
7. Fehilly, C. B., Willadsen, S. M. & Tucker, E. M. Interspecific chimaerism between sheep and goat. *Nature* **307**, 634–636 (1984).
8. Williams, T. J., Munro, R. K. & Shelton, J. N. Production of interspecies chimeric calves by aggregation of *Bos indicus* and *Bos taurus* demi-embryos. *Reprod. Fertil. Dev.* **2**, 385–394 (1990).
9. Rossant, J., Croy, B. A., Chapman, V. M., Siracusa, L. & Clark, D. A. Interspecific chimeras in mammals: a new experimental system. *J. Anim. Sci.* **55**, 1241–1248 (1982).
10. Rossant, J., Mauro, V. M. & Croy, B. A. Importance of trophoblast genotype for survival of interspecific murine chimeras. *J. Embryol. Exp. Morphol.* **69**, 141–149 (1982).
11. Wu, J. & Izpisua Belmonte, J. C. Dynamic pluripotent stem cell states and their applications. *Cell Stem Cell* **17**, 509–525 (2015).
12. Xiang, A. P. *et al.* Extensive contribution of embryonic stem cells to the development of an evolutionarily divergent host. *Hum. Mol. Genet.* **17**, 27–37 (2008).
- The first study reporting PS-cell-derived viable interspecies chimaeras between two distant rodent species.**
13. Han, X. *et al.* Forebrain engraftment by human glial progenitor cells enhances synaptic plasticity and learning in adult mice. *Cell Stem Cell* **12**, 342–353 (2013).
14. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **433**, 519–523 (2008).
- A landmark paper describing the pluripotent ground state and how the ground state culture ultimately enabled derivation of genuine ES cells from rat blastocysts (see also refs 15, 16).**
15. Buehr, M. *et al.* Capture of authentic embryonic stem cells from rat blastocysts. *Cell* **135**, 1287–1298 (2008).
16. Li, P. *et al.* Germline competent embryonic stem cells derived from rat blastocysts. *Cell* **135**, 1299–1310 (2008).
17. Kobayashi, T. *et al.* Generation of rat pancreas in mouse by interspecific blastocyst injection of pluripotent stem cells. *Cell* **142**, 787–799 (2010).
- A landmark paper reporting PS-cell-derived viable mouse–rat interspecies chimaeras and demonstrating the first proof-of-concept of interspecies blastocyst complementation.**
18. Isotani, A., Hatayama, H., Kaseda, K., Ikawa, M. & Okabe, M. Formation of a thymus from rat ES cells in xenogeneic nude mouse–rat ES chimeras. *Genes Cells* **16**, 397–405 (2011).
19. Thomson, J. A. *et al.* Isolation of a primate embryonic stem cell line. *Proc. Natl Acad. Sci. USA* **92**, 7844–7848 (1995).
20. Sasaki, E. *et al.* Establishment of novel embryonic stem cell lines derived from the common marmoset (*Callithrix jacchus*). *Stem Cells* **23**, 1304–1313 (2005).
21. Simerly, C. R. *et al.* Establishment and characterization of baboon embryonic stem cell lines: an Old World Primate model for regeneration and transplantation research. *Stem Cell Res. (Amst.)* **2**, 178–187 (2009).
22. Thomson, J. A. *et al.* Pluripotent cell lines derived from common marmoset (*Callithrix jacchus*) blastocysts. *Biol. Reprod.* **55**, 254–259 (1996).
23. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
24. Reubinoff, B. E., Pera, M. F., Fong, C. Y., Trounson, A. & Bongso, A. Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. *Nat. Biotechnol.* **18**, 399–404 (2000).
25. Tachibana, M. *et al.* Generation of chimeric rhesus monkeys. *Cell* **148**, 285–295 (2012).
26. Simerly, C. *et al.* Interspecies chimera between primate embryonic stem cells and mouse embryos: monkey ESCs engraft into mouse embryos, but not post-implantation fetuses. *Stem Cell Res. (Amst.)* **7**, 28–40 (2011).
27. James, D., Noggle, S. A., Swigut, T. & Brivanlou, A. H. Contribution of human embryonic stem cells to mouse blastocysts. *Dev. Biol.* **295**, 90–102 (2006).
28. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).
29. Boroviak, T., Loos, R., Bertone, P., Smith, A. & Nichols, J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* **16**, 516–528 (2014).
30. Brons, I. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195 (2007).
31. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
32. Huang, Y., Osorno, R., Tsakiridis, A. & Wilson, V. *In vivo* differentiation potential of epiblast stem cells revealed by chimeric embryo formation. *Cell Reports* **2**, 1571–1578 (2012).
33. Kojima, Y. *et al.* The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell* **14**, 107–120 (2014).
34. Wu, J. *et al.* An alternative pluripotent state confers interspecies chimaeric competency. *Nature* **521**, 316–321 (2015).
35. Mascetti, V. L. & Pedersen, R. A. Human–mouse chimerism validates human stem cell pluripotency. *Cell Stem Cell* **18**, 67–72 (2016).
- Refs 34, 35 are two pioneering studies that demonstrated functional engraftment of primed human PS cells into gastrula-stage mouse embryos.**
36. Wu, J. & Izpisua Belmonte, J. C. Stem cells: a renaissance in human biology research. *Cell* **165**, 1572–1585 (2016).
37. Mascetti, V. L. & Pedersen, R. A. Contributions of mammalian chimeras to pluripotent stem cell research. *Cell Stem Cell* **19**, 163–175 (2016).
38. Hanna, J. *et al.* Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc. Natl Acad. Sci. USA* **107**, 9222–9227 (2010).
- A pioneering study demonstrating that a naive human pluripotent state, resembling that of mouse ES cells, could be stabilized in culture.**
39. Chan, Y.-S. *et al.* Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell* **13**, 663–675 (2013).
- Refs 39, 41–43, 46 are examples of the first studies describing different culture conditions for naive human ES cells.**
40. Duggal, G. *et al.* Alternative routes to induce naive pluripotency in human embryonic stem cells. *Stem Cells* **33**, 2686–2698 (2015).
41. Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).
42. Takashima, Y. *et al.* Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269 (2014).
43. Theunissen, T. W. *et al.* Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487 (2014).
44. Theunissen, T. W. *et al.* Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* **19**, 502–515 (2016).
45. Wang, J. *et al.* Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405–409 (2014).
46. Ware, C. B. *et al.* Derivation of naive human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **111**, 4484–4489 (2014).
47. Guo, G. *et al.* Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. *Stem Cell Rep.* **6**, 437–446 (2016).
48. Pera, M. F. In search of naivety. *Cell Stem Cell* **15**, 543–545 (2014).
49. Huang, K., Maruyama, T. & Fan, G. The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. *Cell Stem Cell* **15**, 410–415 (2014).
50. Masaki, H. *et al.* Interspecific *in vitro* assay for the chimera-forming ability of human pluripotent stem cells. *Development* **142**, 3222–3230 (2015).
51. Chen, Y. *et al.* Generation of cynomolgus monkey chimeric fetuses using embryonic stem cells. *Cell Stem Cell* **17**, 116–124 (2015).
52. Fang, R. *et al.* Generation of naive induced pluripotent stem cells from rhesus monkey fibroblasts. *Cell Stem Cell* **15**, 488–496 (2014).
53. Dupin, E. & Sommer, L. Neural crest progenitors and stem cells: from early development to adulthood. *Dev. Biol.* **366**, 83–95 (2012).
54. Jaenisch, R. Mammalian neural crest cells participate in normal embryonic development on microinjection into post-implantation mouse embryos. *Nature* **318**, 181–183 (1985).
55. Huszar, D., Sharpe, A. & Jaenisch, R. Migration and proliferation of cultured neural crest cells in W mutant neural crest chimeras. *Development* **112**, 131–141 (1991).
56. Cohen, M. A. *et al.* Human neural crest cells contribute to coat pigmentation in interspecies chimeras after *in utero* injection into mouse embryos. *Proc. Natl Acad. Sci. USA* **113**, 1570–1575 (2016).
57. Brüstle, O. *et al.* Chimeric brains generated by intraventricular transplantation of fetal human brain cells into embryonic rats. *Nat. Biotechnol.* **16**, 1040–1044 (1998).
58. Keyour, H. M. *et al.* High-yield selection and extraction of two promoter-defined phenotypes of neural stem cells from the fetal human brain. *Nat. Biotechnol.* **19**, 843–850 (2001).
59. Goyama, S., Wunderlich, M. & Mulloy, J. C. Xenograft models for normal and malignant stem cells. *Blood* **125**, 2630–2640 (2015).
60. Zanjani, E. D., Almeida-Porada, G., Livingston, A. G., Flake, A. W. & Ogawa, M. Human bone marrow CD34⁺ cells engraft *in vivo* and undergo multilineage expression that includes giving rise to CD34⁺ cells. *Exp. Hematol.* **26**, 353–360 (1998).
61. Fujiki, Y. *et al.* Successful multilineage engraftment of human cord blood cells in pigs after *in utero* transplantation. *Transplantation* **75**, 916–922 (2003).
62. Muotri, A. R., Nakashima, K., Toni, N., Sandler, V. M. & Gage, F. H. Development of functional human embryonic stem cell-derived neurons in mouse brain. *Proc. Natl Acad. Sci. USA* **102**, 18644–18648 (2005).
63. Chuma, S. *et al.* Spermatogenesis from epiblast and primordial germ cells following transplantation into postnatal mouse testis. *Development* **132**, 117–122 (2005).
64. Leitch, H. G. *et al.* On the fate of primordial germ cells injected into early mouse embryos. *Dev. Biol.* **385**, 155–159 (2014).
65. Clarke, D. L. *et al.* Generalized potential of adult neural stem cells. *Science* **288**, 1660–1663 (2000).
66. Jiang, Y. *et al.* Pluripotency of mesenchymal stem cells derived from adult marrow. *Nature* **418**, 41–49 (2002).
67. Geiger, H., Sick, S., Bonifer, C. & Müller, A. M. Globin gene expression is reprogrammed in chimeras generated by injecting adult hematopoietic stem cells into mouse blastocysts. *Cell* **93**, 1055–1065 (1998).
68. Harder, F., Henschler, R., Junghahn, I., Lamers, M. C. & Müller, A. M. Human hematopoiesis in murine embryos after injecting human cord blood-derived hematopoietic stem cells into murine blastocysts. *Blood* **99**, 719–721 (2002).

69. Wilson, A. & Trumpp, A. Bone-marrow haematopoietic-stem-cell niches. *Nat. Rev. Immunol.* **6**, 93–106 (2006).
70. Brinster, R. L. & Zimmermann, J. W. Spermatogenesis following male germ-cell transplantation. *Proc. Natl Acad. Sci. USA* **91**, 11298–11302 (1994).
71. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
72. Morata, G. & Ripoll, P. Minutes: mutants of *Drosophila* autonomously affecting cell division rate. *Dev. Biol.* **42**, 211–221 (1975).
73. Amoyel, M. & Bach, E. A. Cell competition: how to eliminate your neighbours. *Development* **141**, 988–1000 (2014).
74. Dejosez, M., Ura, H., Brandt, V. L. & Zwaka, T. P. Safeguards for cell cooperation in mouse embryogenesis shown by genome-wide cheater screen. *Science* **341**, 1511–1514 (2013).
75. Masaki, H. *et al.* Inhibition of apoptosis overcomes stage-related compatibility barriers to chimera formation in mouse embryos. *Cell Stem Cell* **19**, 587–592 (2016).
76. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- This landmark study established that somatic cells could be induced to PS cells by defined transcription factors.**
77. Usui, J. *et al.* Generation of kidney from pluripotent stem cells via blastocyst complementation. *Am. J. Pathol.* **180**, 2417–2426 (2012).
78. Matsunari, H. *et al.* Blastocyst complementation generates exogenic pancreas in vivo in apancreatic cloned pigs. *Proc. Natl Acad. Sci. USA* **110**, 4557–4562 (2013).
79. Rashid, T., Kobayashi, T. & Nakauchi, H. Revisiting the flight of Icarus: making human organs from PS cells with large animal chimeras. *Cell Stem Cell* **15**, 406–409 (2014).
80. Kobayashi, T., Kato-Itoh, M. & Nakauchi, H. Targeted organ generation using *Mixl1*-inducible mouse pluripotent stem cells in blastocyst complementation. *Stem Cells Dev.* **24**, 182–189 (2015).
81. Li, Z. *et al.* 3D culture supports long-term expansion of mouse and human nephrogenic progenitors. *Cell Stem Cell* **19**, 516–529 (2016).
82. Kuehn, M. R., Bradley, A., Robertson, E. J. & Evans, M. J. A potential animal model for Lesch–Nyhan syndrome through introduction of HPRT mutations into mice. *Nature* **326**, 295–298 (1987).
83. Sen, S. Brownback. The Human Chimera Prohibition Act of 2005, S.659 (109th Congress, 2005–2006).
84. Greely, H. T., Cho, M. K., Hogle, L. F. & Satz, D. M. Thinking about the human neuron mouse. *Am. J. Bioeth.* **7**, 27–40 (2007).
85. Hyun, I. *et al.* Ethical standards for human-to-animal chimera experiments in stem cell research. *Cell Stem Cell* **1**, 159–163 (2007).
86. Hermerén, G. Ethical considerations in chimera research. *Development* **142**, 3–5 (2015).
87. Hyun, I. From naïve pluripotency to chimeras: a new ethical challenge? *Development* **142**, 6–8 (2015).
88. Robert, J. S. & Baylis, F. Crossing species boundaries. *Am. J. Bioeth.* **3**, 1–13 (2003).
89. Karpowicz, P., Cohen, C. B. & van der Kooy, D. It is ethical to transplant human stem cells into nonhuman embryos. *Nat. Med.* **10**, 331–335 (2004).
90. Karpowicz, P., Cohen, C. B. & van der Kooy, D. Developing human-nonhuman chimeras in human stem cell research: ethical issues and boundaries. *Kennedy Inst. Ethics J.* **15**, 107–134 (2005).
91. Streiffer, R. in *The Stanford Encyclopedia of Philosophy* (ed. Zalta, E. N.) <http://plato.stanford.edu/entries/chimeras/> (2009).
92. Greely, H. T. in *Oxford Handbook of Animal Ethics* (eds Beauchamp, T. & Frey, R.G.) (Oxford Univ. Press, 2011).
93. Johnston, J. & Eliot, C. Chimeras and “human dignity”. *Am. J. Bioeth.* **3**, W6–W8 (2003).
94. Palacios-González, C. Human dignity and the creation of human–nonhuman chimeras. *Med. Health Care Philos.* **18**, 487–499 (2015).
95. United States National Institutes of Health. Special Notice: Moratorium on Certain Fetal Tissue Research. *NIH Guide for Grants and Contracts*. (Special notice, 8 May 1988).
96. United States National Institutes of Health Revitalization Act of 1993 *Public Law 103-43* (1993).
97. *Sherley v. Sebelius*, 689 F.3d 776 (D.C. Circuit 2012).
98. Committee on Guidelines for Human Embryonic Stem Cell Research. *Guidelines for Human Embryonic Stem Cell Research* (National Academies Press, 2005).
99. International Society for Stem Cell Research. *Guidelines for the Conduct of Human Embryonic Stem Cell Research* <http://www.isscr.org/home/publications/guide-clintrans> (2006).
100. International Society for Stem Cell Research. *Guidelines for Stem Cell Science and Clinical Translation* <http://www.isscr.org/home/publications/2016-guidelines> (2016).
101. Academy of Medical Sciences. *Animals Containing Human Materials* https://issuu.com/acmedsci/docs/animals_containing_human_material. (2011).
102. Greene, M. *et al.* Ethics: moral issues of human-non-human primate neural grafting. *Science* **309**, 385–386 (2005).
103. Palacios-González, C. Ethical aspects of creating human-nonhuman chimeras capable of human gamete production and human pregnancy. *Monash Bioeth. Rev.* **33**, 181–202 (2015).
104. United States National Institutes of Health. *NIH Research Involving Introduction of Human Pluripotent Cells into Non-Human Vertebrate Animal Pre-Gastrulation Embryos*. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-158.html> (2015).
105. United States National Institutes of Health. *Next Steps on Research Using Animal Embryos Containing Human Cells*. <http://osp.od.nih.gov/under-the-poliscope/2016/08/next-steps-research-using-animal-embryos-containing-human-cells> (2016).
106. Reardon, S. US agency to lift ban on funding human–animal hybrids. *Nature* **536**, 135 (2016).

Acknowledgements We would like to thank all our laboratory members and collaborators, including the teams of J. M. Campistol, P. Guillen, E. Martinez and P. Ross for their comments and dedicated work that have greatly contributed to the ideas presented here. H.N. was supported by the California Institute of Regenerative Medicine (CIRM), Japan Agency for Medical Research and Development (AMED) and Takeda Pharmaceuticals International, Inc. J.C.I.B. was supported by the G. Harold and Leila Y. Mathers Charitable Foundation, The Moxie Foundation, Fundación Dr. Pedro Guillen and UCAM. R.J. was supported by grants from the NIH (HD 045022, R37-CA084198, 1R01NS088538-01).

Author Contributions J.W., H.G., R.J., H.N., J.R. and J.C.I.B. conceived the study and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C.I.B. (belmonte@salk.edu).

Reviewer Information Nature thanks S. Goldman, I. Hyun and M. A. Surani for their contribution to the peer review of this work.

Organization and functions of mGlu and GABA_B receptor complexes

Jean-Philippe Pin^{1,2} & Bernhard Bettler³

The neurotransmitters glutamate and γ -aminobutyric acid (GABA) transmit synaptic signals by activating fast-acting ligand-gated ion channels and more slowly acting G-protein-coupled receptors (GPCRs). The GPCRs for these neurotransmitters, metabotropic glutamate (mGlu) and GABA_B receptors, are atypical GPCRs with a large extracellular domain and a mandatory dimeric structure. Recent studies have revealed how these receptors are activated through multiple allosteric interactions between subunit domains. It emerges that the molecular complexity of these receptors is further increased through association with trafficking, effector and regulatory proteins. The structure and composition of these receptors present opportunities for therapeutic intervention in mental health and neurological disorders.

The main excitatory and inhibitory neurotransmitters in the brain, glutamate and GABA, produce fast synaptic responses (<10 ms) by activating ligand-gated channels. Additionally, they influence neuronal and glial processes over a slower timescale (sub-seconds to minutes) by activating mGlu receptors (mGluRs) and GABA_B receptors (GABA_BRs). These receptors activate intracellular heterotrimeric G proteins that regulate enzymes and ion channels influencing neuronal firing, plasticity processes and network activity (reviewed in refs 1, 2). Owing to their modulatory effects on neuronal excitability, mGluRs and GABA_BRs are widely considered to be attractive drug targets for neurological and mental health disorders. GABA_B agonists are in use to treat muscle spasms, alcohol dependency and cataplexy associated with narcolepsy³. Proof-of-concept clinical studies have validated mGlu ligands for the treatment of anxiety, schizophrenia, Parkinson's disease, depression, drug dependency, autism spectrum disorders, obsessive-compulsive disorders, Rhetts syndrome and gastro-oesophageal reflux disease^{4,5}. However, adverse side effects (mainly dizziness, nausea, epileptiform activity, sedation and motor impairment) limit the use of GABA_B compounds and hinder development of mGlu drugs. Recent biophysical and structural studies have greatly improved our understanding of the structural bases of mGluR and GABA_BR activation and modulation. This provides a foundation for developing drugs with a tailored mode of action and possibly fewer side effects.

GABA_BRs and mGluRs are class C GPCRs that also include Ca²⁺-sensing receptors, sweet and umami (the savoury taste of glutamate in humans) taste receptors. Class C GPCRs are structurally more complex than other GPCRs. They form constitutive dimers and each subunit is made of a large extracellular Venus flytrap domain (VFTD) that binds agonists and links to a G-protein-activating seven-transmembrane domain (7TMD)⁶. For the GABA_BRs, three principal subunits, GABA_{B1a}, GABA_{B1b} and GABA_{B2}, constitute two fully functional heterodimeric GABA_{B(1a,2)} and GABA_{B(1b,2)} receptors coupling to G_{i/o}-type G proteins^{1,7} (Fig. 1a). Compared to GABA_{B1b}, GABA_{B1a} contains two extracellular sushi domains at its N terminus that traffic GABA_{B(1a,2)} receptors to axon terminals^{1,8}. As with mGluRs, the eight subunits subdivide into three groups: group I (mGlu₁ and mGlu₅) are mostly postsynaptic receptors that are coupled to G_q proteins, whereas group II (mGlu₂ and mGlu₃), and group III (mGlu₄, mGlu₆, mGlu₇ and mGlu₈) are mostly presynaptic receptors that are coupled to G_{i/o} proteins². Subunits of mGluRs typically

constitute homodimeric receptors but recent studies revealed the existence of heterodimeric receptors made of specific subunit combinations⁹, which potentially generates 11 additional receptors (Fig. 1b).

Here we describe biophysical and structural studies that revealed the mode of activation of dimeric GABA_BRs and mGluRs. Proteomic approaches, mainly with the GABA_BRs, revealed how receptor-associated proteins work together with principal receptor subunits and G proteins to produce a variety of molecularly and functionally distinct multiprotein-receptor complexes. Through proteomic approaches, novel signalling pathways for GABA_BRs have also been revealed, which has implications for our understanding of receptor function in the nervous system and also presents opportunities for drug discovery. The progress made with GABA_BRs and mGluRs may also help our understanding of other GPCRs that form heteromers and associate with partner proteins.

Receptor oligomers and signalling efficacy

GABA_BRs have provided the first indisputable example for obligate GPCR heterodimers because two structurally related subunits, GABA_{B1} and GABA_{B2}, are necessary for the formation of a functional receptor in transfected cells⁷. The heteromeric structure of GABA_BRs was validated *in vivo* using genetic¹, biophysical¹⁰ and biochemical approaches^{11,12}. It is well-established that GABA_{B1} contains the agonist-binding site, whereas GABA_{B2} couples to the G protein¹³ (Fig. 1a). GABA_{B2} also allosterically increases agonist affinity at GABA_{B1} (ref. 7).

Assembly of multiple subunits is associated with an increased risk of protein misfolding¹⁴. GABA_BRs therefore pass through a quality control system that prevents unfolded and unassembled subunits from exiting the endoplasmic reticulum. Quality control relies on the arginine-based endoplasmic-reticulum-retention motif RSRR (Fig. 1a) that keeps GABA_{B1} in the endoplasmic reticulum by binding to the prenylated Rab acceptor family 2 (PRAF2) protein¹⁵. The retention motif is located distal to a coiled-coil domain in GABA_{B1} that interacts with a coiled-coil domain in GABA_{B2} (ref. 16) (Fig. 1a). Coiled-coil interaction competitively displaces PRAF2 from GABA_{B1} by rendering the RSRR motif inaccessible, thereby releasing the heterodimeric receptor complex from the endoplasmic reticulum.

mGlu subunits also form dimers that are constitutively stabilized by an inter-subunit disulfide bridge^{9,17–19} (Figs 1b, 2a). A dimeric assembly is necessary for agonist activation, as shown by analysis of purified

¹Centre National de la Recherche Scientifique UMR 5203, Institut de Génétique Fonctionnelle, Université de Montpellier, F-34094 Montpellier, France. ²INSERM, U1191, F-34094 Montpellier, France. ³Department of Biomedicine, Pharmazentrum, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland.

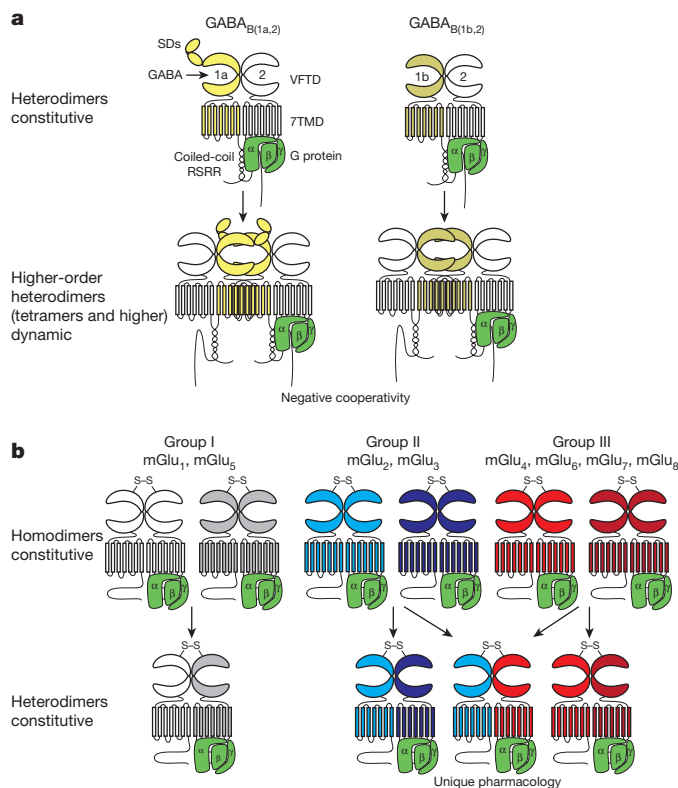


Figure 1 | Oligomerization of GABA_BRs and mGluRs. **a**, The minimal functional units of GABA_BRs are constitutive heterodimers of GABA_{B1a} (1a) or GABA_{B1b} (1b) subunits with the GABA_{B2} (2) subunit docked to the heterotrimeric G protein ($\alpha\beta\gamma$)^{1,7}. The heterodimer is stabilized by coiled-coil interactions. The GABA_{B1a} subunit contains two N-terminal sushi domains (SDs) that mediate axonal localization of the receptor, and the GABA_{B1b} subunit mediates dendritic localization¹. The GABA binding site is confined to the VFTD of the GABA_{B1} subunit. Heterodimers may further dynamically self-assemble into tetramers (dimers of heterodimers) or even higher-order oligomers, at least partly through interactions between the VFTD of the GABA_{B1} subunits^{10,11,23,24}. Tetramers exhibit negative cooperativity and activate the G protein less efficiently¹⁰. It is unknown whether tetramers can activate more than one G protein. **b**, mGluRs assemble into constitutive homodimers stabilized by intermolecular disulfide bridges (S–S) in their VFTDs^{9,17–19}. Glutamate binding to the VFTDs of both subunits is required for efficient signalling¹⁸, but only one of the two subunits activates the G protein at a time⁴⁸. mGluRs can form heterodimers within their classification groups and between group-II and -III (ref. 9) in transfected cells. mGlu₂–mGlu₄ heterodimers have been identified in native tissue based on unique pharmacological properties²¹. mGlu₁–mGlu₅ (ref. 22) and mGlu₇–mGlu₈ (ref. 99) are likely candidates for *in vivo* heteromers on the basis of proteomic and expression data, respectively.

monomers and dimers reconstituted into lipid nanodiscs²⁰. In contrast to GABA_BRs, mGluRs were for a long time thought to exclusively form homodimers, owing to largely non-overlapping expression patterns in the brain. However, quantitative fluorescence resonance energy transfer (FRET) experiments revealed the existence of mGlu heterodimers in transfected cells, composed of either group I, or group II and group III subunits⁹ (Fig. 1b). Experimental support for the existence of heterodimeric mGluRs in native tissue is still scarce, but pharmacological and proteomic evidence supports the existence of mGlu₂–mGlu₄ and mGlu₁–mGlu₅ heterodimers *in vivo*^{21,22}.

There is a consensus in the field that mGluRs are limited to strict dimers and do not form higher-order complexes^{9,17,18}, although this remains to be confirmed with native receptors. By contrast, several studies show that GABA_BRs form tetramers or larger complexes through dynamic self-assembly of heterodimeric units^{10,23,24} (Fig. 1a). Single-molecule tracking studies in living cells show that most GABA_BRs form tetramers

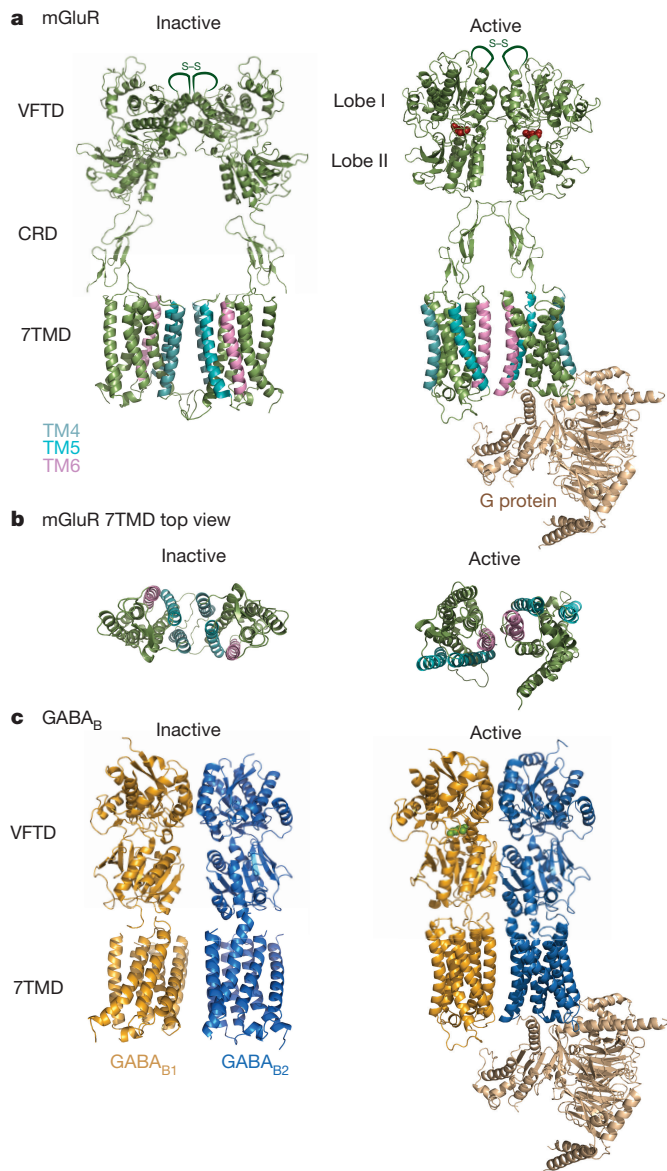


Figure 2 | Structural models of dimeric mGluRs and GABA_BRs in the inactive and active states. **a**, The inactive mGluR model is based on the dimeric mGlu₃ VFTDs (Protein Data Bank (PDB), 3SM9), to which the cysteine-rich domains (CRD) were added according to the extracellular domain of mGlu₃ (PDB, 2E4U)²⁸. The active dimeric extracellular domain is based on the structure of mGlu₃ bound to glutamate with the relative orientation of the extracellular domains of the symmetric mGlu₁ VFTD dimer²⁷. The flexible loops are covalently linked by a disulfide bond (S–S). The inactive dimeric 7TMDs are based on the association of the mGlu₅ 7TMD structure (PDB, 4OO9)³³ through TM4 and TM5 (depicted in dark green and blue)⁴⁹. The active form model with associated nucleotide-free G protein was based on the β 2-adrenergic-receptor–G-protein complex (PDB, 3SN6) associated with an inactive 7TMD of mGlu₁ through TM6 (depicted in magenta)⁴⁹. Note that it is unknown how the extracellular domain interacts with the 7TMD. **b**, Top view of the 7TMD dimers in their proposed inactive and active states. **c**, The inactive and active models of the GABA_BR (GABA_{B1} in yellow, GABA_{B2} in blue) correspond to an apposition of the solved structures of the heterodimeric VFTDs with the GABA_{B1} VFTD bound to antagonist (PDB, 4MR7) or GABA (PDB, 4SM3)²⁹, with the model of the 7TMD dimers generated similar to that of the mGluR. It is, however, neither known whether the GABA_BR 7TMDs are associated similarly to those of mGluRs, nor whether they undergo a similar change in conformation during activation.

at low receptor density and larger complexes at high density²⁴. The existence of higher-order GABA_BR complexes is supported by analysis of solubilized native receptors and by freeze-fracture electron microscopy

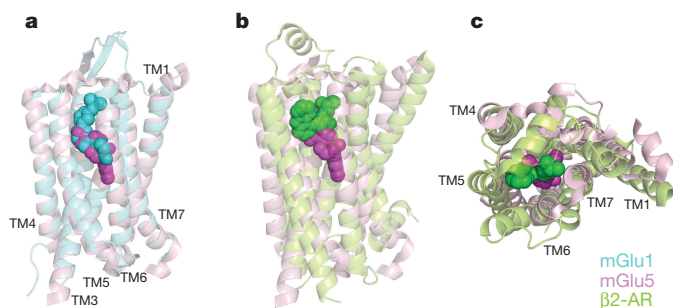


Figure 3 | Ribbon views of the 7TMDs of mGlu₁, mGlu₅ and the ̢₂-adrenergic receptor (̢₂-AR), with bound FITM, mavoglurant and (S)-carazolol. a, Superposition of the mGlu₁ (light blue, PDB, 4OR2) and mGlu₅ (light magenta, PDB, 4OO9) 7TMD structures, illustrating the different binding sites for FITM (to mGlu₁, blue) and mavoglurant (to mGlu₅, magenta). **b, c**, Superposition of the mGlu₅ and ̢₂-AR (light green, PDB, 2RH1) structures from the side (**b**) and from the top (**c**), highlighting the deeper location of mavoglurant compared to (S)-carazolol (to ̢₂-AR, green).

in hippocampal neurons¹¹. Higher-order GABA_BR complexes are also revealed in FRET studies with tagged or antibody-labelled subunits in neurons¹⁰. The two GABA_{B1} subunits in tetramers contact each other, and the four VFTDs are probably organized similarly to the VFTDs of α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors²⁵. GABA_BR tetramers are stable, with no exchange of heterodimers within pre-existing tetramers¹⁰. Tetrameric assembly of GABA_BRs restricts G-protein activation²³, but may allow interaction of the G protein with other receptor-interacting proteins. Clearly, more work is necessary to determine the functional relevance of higher-order GABA_BR complexes.

Active and inactive receptor structures

The X-ray structures of isolated mGlu VFTDs revealed that agonists stabilize the closed conformation of the VFTDs^{26–28}. This was later also observed with the GABA_{B1} VFTD²⁹. In line with the structural data, mutations that prevent or stabilize VFTD closure lead to inactive or constitutively active receptors, respectively^{30–32}. VFTD closure therefore constitutes a first step in the mGluR and GABA_BR activation process (Fig. 2).

Recently, the 7TMD structures of mGlu₁ and mGlu₅ were solved^{33–35} (Figs 2, 3). The 7TMD organization is similar to that of class A GPCRs, despite considerable sequence divergences. The 7TMD is arranged around the third transmembrane domain (TM3), which is centrally located at the extracellular side of the membrane (Fig. 3). Compared to class A GPCRs, the helices are more compact, with an inward shift of TM5 that renders the surface area composed of TM4, TM5 and TM6 quite flat. The 7TMD structures were solved in the presence of the negative allosteric modulators (NAMs) FITM (for mGlu₁) and mavoglurant (for mGlu₅), the latter compound is under clinical evaluation for levodopa-induced dyskinesia and obsessive–compulsive disorder. These NAMs stabilize an inactive 7TMD conformation and interact in a cavity located centrally between TM3, TM5, TM6 and TM7. The FITM-binding site corresponds to the binding site of small ligands in many class A GPCRs, located above the tryptophan residue in TM6 that is conserved in both class A and C GPCRs. Mavoglurant binds deeper in the 7TMD core, entering a narrow cavity that can adapt several mGlu₅-specific ligands. Distinct binding pockets for NAMs in mGlu₁ and mGlu₅ reveal that multiple sites in class C 7TMDs can bind allosteric modulators. Mutagenesis and modelling also demonstrated distinct allosteric sites and binding modes within the 7TMD of mGlu₅ and mGlu₄ (refs 36,37). These structural observations need to be related to the differing pharmacological properties of allosteric modulators. For example, some NAMs are inverse agonists³⁸ while others are partial inhibitors³⁹. Likewise, some but not all positive allosteric modulators (PAMs) display strong agonist activity³⁷, whereas others bias intracellular signalling to specific pathways⁴⁰. Structural considerations may therefore guide the development of compounds with tailored properties.

Dimerization of class C GPCRs is required for agonist activation of the 7TMDs but not for G-protein coupling²⁰. Isolated 7TMDs of monomeric mGluRs fully activate the G protein upon binding of compounds that act as PAMs at the full-length receptor dimer²⁰. It is assumed that the 7TMDs of class C GPCRs undergo a similar conformational change as class A GPCRs when activating the G protein. Similar to most class A GPCRs, an ionic lock is found at the intracellular side of mGluRs^{33,34} and GABA_BRs⁴¹. This lock is different from class A GPCRs and involves glutamate or aspartate (TM6) and lysine (TM3) residues that are highly conserved in class C GPCRs. The ionic lock stabilizes the inactive closed 7TMD conformation on the cytoplasmic side. Accordingly, mutation of the ionic lock in GABA_BRs leads to increased agonist affinity⁴¹, whereas mutation in mGlu₅ leads to increased constitutive activity³³.

Allosteric interactions and receptor activation

No X-ray structure of a full-length class C GPCR is available yet. Nevertheless, much information on how the closure of the VFTD activates the 7TMDs is available. In mGluRs, the VFTDs associate through a hydrophobic interface in the N-terminal lobes (lobe I). The dimeric structure is stabilized by an inter-subunit disulfide bond connecting two flexible loops at the top of the VFTDs (Fig. 2a). X-ray structures suggested a major reorientation of the two domains when bound to an agonist versus an antagonist, with lobe II on each C terminus being closer together in the agonist bound form²⁶ (Fig. 2a). This is consistent with FRET studies with full-length receptors expressed at the cell surface of transfected cells and with cross-linking experiments locking the VFTD dimers in either the active or inactive orientation⁴². Single-molecule FRET studies revealed that isolated VFTD dimers oscillate between active and inactive orientations in the sub-millisecond time scale⁴³, consistent with additional structural data showing that inactive and active dimer orientations also occur when bound to agonists and antagonists, respectively^{28,42}. Fast conformational dynamics were not seen when analysing full-length receptors with single-molecule FRET⁴⁴, possibly owing to a lower time resolution of the technique used or because the 7TMDs limit VFTD dynamics. Overall, the data support that ligands act on pre-existing states (selection model of allostery) rather than by inducing a specific conformation of the receptor (conformational induction model). In addition, single-molecule analysis of full-length dimers revealed an intermediate conformational state that may be essential for the transition from the inactive to the active VFTD dimer orientation⁴⁴. Single-molecule FRET studies revealed that agonists act by controlling the equilibrium constant between different states, leading to a higher proportion of receptors in the active state⁴³, which is in line with the selection model of allostery. Accordingly, a partial agonist was found to be less efficient in stabilizing the active state than the full agonist glutamate^{43,44}.

In the mGluR dimer, closure of one VFTD in response to glutamate is sufficient to activate the G protein, but closure of both VFTDs is required for maximal activation¹⁸. It appears that a smaller distance between the VFTD C termini increases the coupling efficacy of the receptor dimer. The movement of the VFTD C termini is amplified through the cysteine-rich domain (Fig. 2a), a rigid extension at the bottom of the VFTD, as revealed by the X-ray structure of the entire N-terminal extracellular domain of mGlu₅ (ref. 28). The link between the VFTD and the cysteine-rich domain is rigidified by a disulfide bridge interconnecting these two domains, which is essential for communication of the VFTDs with the 7TMDs⁴⁵. The dimeric extracellular domains of the receptor principally control the distance between the TM1 domains, and thereby the relative position of the 7TMDs. Movements of both 7TMDs in mGluRs were confirmed through FRET experiments⁴⁶. Detailed analysis revealed a two-step process: first a relative movement between the 7TMDs with a half-life ($t_{1/2}$) of approximately 25 ms, and a second conformational change in only one 7TMD with a $t_{1/2}$ of approximately 10 ms⁴⁷. In agreement with this process, previous functional studies indicate that the 7TMD dimer of homodimeric receptors is asymmetric, with only one 7TMD active at a time⁴⁸. Analysing spontaneous crosslinking events of cysteine mutations in the 7TMDs revealed that the two subunits mainly interact through

TM4 or TM5 in the inactive conformation, whereas crosslinking of both TM6 is favoured in the active conformation⁴⁹ (Fig. 2b). Accordingly, TM4 or TM5 crosslinking prevents agonist activation of the receptor, whereas TM6–TM6 crosslinking generates a constitutively active receptor⁴⁹. In summary, the current model of mGluR activation proposes that the closure of the VFTDs stabilizes a different relative orientation of the VFTDs that brings the TM1 of the two 7TMDs closer together (Fig. 2a). The resulting rotation of one 7TMD relative to the other allows the two TM6 to contact each other, which results in the activation of one 7TMD (Fig. 2b). A change in the interaction between the linker connecting TM1 to the cysteine-rich domain and the extracellular loop 2, as observed in the mGlu₁ 7TMD structure³⁴, may also control the 7TMD conformation. Only one G protein can bind in the TM6 dimer of mGluRs, thus stabilizing only one 7TMD in its active conformation⁴⁹. Alternatively, steric hindrance within the active 7TMD dimer interface prevents movement of both TM6 domains, thus enabling only one subunit to be active at a time. These observations provide a structural hypothesis for the asymmetric activation of both class C and class A GPCR dimers⁵⁰.

GABA_BRs and mGluRs share similar VFTDs and 7TMDs, however their activation mechanisms differ. GABA_BRs lack a cysteine-rich domain and only one agonist per dimer is sufficient for full activation²⁹ (Fig. 2c). Moreover, the GABA_{B2} VFTD does not close during receptor activation³². The relative movement between the VFTDs that is sufficient for GABA_BR activation is much smaller than in mGluRs, as revealed by the crystal structures of the GABA_B VFTD heterodimer with and without bound agonist²⁹ (Fig. 2c). GABA_{B1} VFTD closure in response to agonist binding leads to activation of the GABA_{B2} 7TMD in two synergistic ways: (1) through a conformational change in the GABA_{B1} 7TMD activating the GABA_{B2} 7TMD; and (2) through the GABA_{B2} VFTD stabilizing the closed form of the GABA_{B1} VFTD and activating the GABA_{B2} 7TMD⁵¹. The compactness of GABA_BRs therefore enables activation of one 7TMD per dimer through a direct allosteric coupling between associated domains that requires minimal relative movements between domains.

The models of mGluR and GABA_BR activation suggest a mode of action for allosteric modulators acting in the 7TMDs. NAMs probably prevent the 7TMDs of the two subunits of mGluRs⁵² and the GABA_{B2} subunit of GABA_BRs⁵³ from reaching an active conformation. By contrast, PAMs probably stabilize an active conformation of at least one 7TMD upon agonist activation in the VFTDs. This is well illustrated by the agonist activity of PAMs at 7TMDs that lack the extracellular domain⁵⁴. This implies that the extracellular domain prevents PAMs from directly activating the 7TMD. However, some PAMs display agonist activity and activate the receptor without agonist binding in the VFTD. The reason for this agonist activity is unclear but may relate to the different binding modes of PAMs in the 7TMDs, as proposed for mGlu₄ (ref. 37).

Receptor-associated proteins

GPCR kinases (GRKs) and arrestins usually regulate internalization, desensitization and intracellular signalling of GPCRs⁵⁵. However, mGluR and GABA_BR responses desensitize in the absence of GRK enzymatic activity^{56,57} and arrestins are not required for receptor internalization^{56,58,59}. These findings gave rise to efforts to identify other proteins that regulate GABA_BR and mGluR signalling. Numerous receptor-associated proteins were identified using yeast two-hybrid screens and related approaches (reviewed in refs 60–64). Unfortunately, many of these proteins were not validated for their interaction with native receptors. Here we discuss the ensemble of proteins that directly or indirectly interact with the native receptors (referred to as the GABA_BR and mGluR interactomes; Tables 1, 2).

The GABA_BR interactome

Comprehensive proteomic approaches identified approximately 30 high-confidence components of GABA_BRs in the rodent brain^{11,12,65}. Surprisingly, most of these components have neither been linked to GPCR signalling nor were they isolated in previous yeast two-hybrid screens, which may miss interacting proteins with multiple contact points.

Conversely, proteins selectively identified in yeast two-hybrid screens may represent transient interactors or proteins of low abundance in native receptor complexes¹². The components of native GABA_BRs generally exhibit differential spatiotemporal expression patterns in the brain, thus limiting the number of possible receptor constituents in individual cells and supporting a modular composition of receptors¹².

Highly enriched in native GABA_BR complexes are the four cytosolic K⁺-channel tetramerization-domain (KCTD) proteins KCTD8, KCTD12, KCTD12b and KCTD16 (referred to as the KCTDs)^{11,12} (Table 1). The KCTDs form tetramers that constitutively bind to the GABA_{B2} C-terminal domain¹¹. KCTD tetramers additionally bind to G protein $\beta\gamma$ subunits and tether the G protein to the receptor⁶⁵. Pre-assembly of the G protein to the receptor renders signalling independent of random collision coupling⁶⁶ and accelerates G-protein signalling⁶⁵. KCTD12 and KCTD12b also induce a fast (within seconds) desensitization of receptor-induced K⁺ currents by occupying an activity-dependent binding site on the G $\beta\gamma$ subunits⁶⁵. Mechanistically, activity-dependent binding of KCTD12 uncouples the G $\beta\gamma$ subunits from effector K⁺ channels, which desensitizes K⁺ currents. KCTD8 selectively reduces tonic G-protein activation by the receptor and thereby decreases agonist-independent signalling⁶⁷, whereas KCTD16 serves as an adaptor for effector channels and other associated proteins¹² (Table 1). The KCTDs are widespread GABA_BR components that qualify as auxiliary receptor subunits. Reverse-affinity purification experiments support that the KCTDs exclusively bind to GABA_BRs and not to other GPCRs^{12,65}. However, the KCTDs may influence other GPCRs through their binding to G $\beta\gamma$ subunits⁶⁵. Principal GABA_{B1} and GABA_{B2} subunits and auxiliary KCTD subunits, together with the G protein, are now viewed as the core building blocks of GABA_BRs (Table 1).

Peripheral building blocks bind to core building blocks and generate receptor complexes with unique functions and localization. Peripheral interactors comprise effector channels, elements of the presynaptic release machinery, sushi-domain-interacting proteins, neuronal adhesion molecules and proteins that regulate G-protein signalling¹² (Table 1). Several proteins bind to the core building blocks transiently and regulate receptor surface expression (Table 1). The proteomic data support that peripheral components interact with the receptor either through the C-terminal domains of GABA_{B1} and/or GABA_{B2} or through KCTD16 or the sushi domains of GABA_{B1a}.

Proteomic analysis revealed unanticipated GABA_BR effector channels. KCTD16 was shown to scaffold hyperpolarization-activated cyclic-nucleotide gated (HCN) channels at the receptor, which promotes HCN channel activation through receptor-induced hyperpolarizing Kir3 currents¹². Preventing the GABA_BR–HCN interaction in KCTD16 knockout mice interferes with HCN channel activation and influences the shape and propagation of synaptic potentials¹². Moreover, GABA_{B1} associates in a ligand-independent manner with transient receptor-potential vanilloid 1 (TRPV1) channels in dorsal root ganglia⁶⁸. Pharmacological activation of GABA_{B1} inhibits TRPV1 sensitization in inflammatory pain conditions, thus making GABA_{B1} a drug target for chronic pain indications⁶⁸.

Most of the receptor-associated proteins identified in native tissue associate with GABA_{B1a}- and GABA_{B1b}-type receptors¹². Notable exceptions are the adherence-junction associated protein-1 (AJAP1), amyloid-precursor protein (APP), amyloid-precursor-like protein-2 (APLP2) and PILR-associated neural protein. These four proteins bind to the sushi domains of GABA_{B1a} (Table 1) and are expected to be involved in subcellular localization of GABA_BRs¹. APP simultaneously binds to the sushi domains and to the integral-membrane proteins 2B/C¹² (ITM2B/C), the c-Jun N-terminal kinase (JNK)-interacting protein-3 (JIP-3) and calyntenin-3, which therefore are secondary interactors of GABA_BRs. As JIP-3 and calyntenin-3 bind kinesin-1, they may link GABA_{B1a} through APP to the axonal motor^{12,69–71}.

Several proteins of the forward trafficking pathway enrich in purified receptor complexes, including reticulocalbin-2, calnexin and 14-3-3 family members¹² (Table 1). 14-3-3 proteins are proposed to dissociate and impair GABA_BR heterodimers in spinal cord neurons and to underlie chronic pain conditions⁷². 14-3-3 proteins also uncouple GABA_BRs from Kir3 channels to mediate anti-depressant effects⁷³.

Table 1 | GABA_BR interactome

Receptor components	Site of interaction	GABA _B R-related function	Refs
Core components			
GABA _{B1a}	GABA _{B2}	Agonist binding, presynaptic localization	87
GABA _{B1b}	GABA _{B2}	Agonist binding, postsynaptic localization	87
GABA _{B2}	GABA _{B1a} , GABA _{B1b}	Obligate heteromer with GABA _{B1} subunits	11, 13
G proteins	G-protein subunits	G protein docking	65
	KCTD8, -12, -12b, -16	Interaction with auxiliary KCTD subunits	
	GABA _{B2}	Signalling to effector channels and adenylyl cyclase	
KCTD8	KCTD8, -12, -12b, -16	Interaction with auxiliary KCTD subunits	11, 65, 67
	CT GABA _{B2}	Accelerates G protein signalling	
KCTD12	G protein $\beta\gamma$ subunits	Reduces constitutive activity	11, 65
	CT GABA _{B2}	Accelerates G protein signalling	
KCTD12b	G protein $\beta\gamma$ subunits	Interferes with G $\beta\gamma$ signalling	11, 65
	CT GABA _{B2}	Promotes K ⁺ -current desensitization	
	GABA _{B2}	Accelerates G protein signalling	
KCTD16	G protein $\beta\gamma$ subunits	Interferes with G $\beta\gamma$ signalling	11, 12, 65
	CT GABA _{B2}	Promotes K ⁺ -current desensitization	
	GABA _{B2}	Accelerates G protein signalling	
KCTD16	G protein $\beta\gamma$ subunits	Scaffolds HCN channels at GABA _B R	
Peripheral components with known site of interaction			
AJAP1	SDs GABA _{B1a}	ND, receptor localization?	12
APP	SDs GABA _{B1a}	ND, receptor localization?	12
APLP2	SDs GABA _{B1a}	ND, receptor localization?	12
PILR-associated neural protein	SDs GABA _{B1a}	ND, receptor localization?	12
CHOP†	CT GABA _{B1}	Intracellular retention	61
CamKII*	CT GABA _{B2}		
	CT GABA _{B1}	Promotes receptor internalization	
COPI†	CT GABA _{B1}	Intracellular retention	88
CREB†	CT GABA _{B1}	Mediates GABA _B effects on gene expression	61
GABA _A γ2S*	CT GABA _{B2}		60, 61
	GABA _{B1}	Promotes surface expression of GABA _{B1}	
GISP*	GABA _{B1}	Enhances agonist-induced receptor internalization	89
	CT GABA _{B1}	Enhances receptor surface expression	
msec-7-1*	CT GABA _{B1}	Enhances receptor surface expression	12, 60, 61
PRAF2†	CT GABA _{B1}	Intracellular retention of GABA _{B1}	15
Syntaxin-1	CT GABA _{B1}	Regulates probability of synaptic vesicle fusion	90
TRPV1	GABA _{B1}	Effector channel	68
Marlin1*	CT GABA _{B1}	ND, receptor trafficking?	12, 60, 61
14-3-3 proteins	CT GABA _{B1}	Impair GABA _B R heterodimers	72, 73
	KCTD8/16	Uncouple GABA _B Rs from Kir3 channels	
Mupp1*	CT GABA _{B2}	Enhances GABA _B signalling	12, 60
NSF*	CT GABA _{B2}	Promotes desensitization	12, 60
HCN2	KCTD16	Effector channel	12
Ca ²⁺ channel (Ca _v 2.2)	KCTD16	Effector channel	12
GINIP*	G protein $\alpha_{i/o}$ subunits	Prolongs G protein signalling	12, 91
Calsynt-3‡	APP/APLP2	ND	12
ITM2B/C‡	APP/APLP2	ND	12
JIP-3‡	APP	ND	12
Peripheral components with unknown site of interaction			
Calnexin	ND	ND	12
DPP-6/10	ND	ND	12
Gβ5	ND	Binds to RGS7 and fastens K ⁺ -current deactivation	92
GRK4†	ND	Desensitization	57
Neurologin-3	ND	ND	12
Reticulocalbin-2	ND	ND	12
Synaptotagmin-11	ND	ND	12

The principal subunits GABA_{B1a}, GABA_{B1b} and GABA_{B2} assemble into two heterodimeric receptors composed of a GABA_{B1} and a GABA_{B2} subunit⁷. GABA_{B1a} differs from GABA_{B1b} by a pair of extracellular sushi-domains (SDs) that function as axonal trafficking signals⁸⁷ and bind AJAP1, APP, APLP2 and PILR-associated protein. Auxiliary KCTD8, -12, -12b and -16 proteins bind to both the C terminus (CT) of GABA_{B2} and the G protein $\beta\gamma$ subunits, which preassembles the G protein at the receptor and regulates kinetic properties of the receptor response⁶⁵. KCTD8 and KCTD16 also scaffold effector channels and 14-3-3 proteins at the receptor¹². Principal and auxiliary receptor subunits, together with the G protein, constitute the core components of the receptor to which several peripheral components can bind to. Secondary interactors (indicated with †) do not directly bind to core receptor components. Assembly with peripheral components generates a variety of molecularly distinct receptor complexes that regulate G-protein signalling, receptor localization, surface trafficking, channel gating or neurotransmitter release. Most proteins identified in native receptor complexes bind to receptors assembled with GABA_{B1a} and GABA_{B1b}. Exceptions are proteins that directly bind to the sushi domains of GABA_{B1a} and secondary interactors binding to APP or APLP2 (ref. 65).

*Proteins that are present in purified native receptor complexes but do not fulfill stringent significance criteria¹², possibly because they transiently interact with the receptor.

†Proteins that were not found in purified native receptor complexes but for which functional effects on GABA_BR signalling were reported. ND, not determined.

The mGluR interactome

Unbiased comprehensive affinity purifications of native mGluRs are missing. Most proteins that interact with mGluRs have been identified using yeast two-hybrid screens or affinity purifications with synthetic peptides (reviewed in refs 63, 64). Interacting proteins include scaffold

proteins (for example, Homer, GRIP-1, NHERF-2, PICK-1, tamalin), ion channels (for example, Ca²⁺ channels, NMDA receptors), signalling molecules (for example, calmodulin, GRK2, ephrin-B2) and elements of the cytoskeleton (for example, filamin-1, microtubule-associated protein-1B). Currently, G-protein subunits, Ca²⁺ and TRP channels

Table 2 | Group I mGluR interactome

Receptor components	Site of interaction	mGluR _{1/5} -related function	Refs
Core components			
mGlu ₁		Principal subunit forming receptor dimers	9, 22
mGlu ₅		Principal subunit forming receptor dimers	9, 22
G proteins	Cavity between IL2 and IL3	Signalling to effectors	48
Homer1	CT 1a, 5a/b	Regulates localization and signalling to effectors	63, 93
Homer1a	CT 1a, 5a/b	Promotes constitutive activity	38, 63, 93
Homer2	CT 1a, 5a/b	Regulates localization and signalling to effectors	63, 93
Homer3	CT 1a, 5a/b	Regulates localization and signalling to effectors	38, 63, 93
		Reduces constitutive activity	
Peripheral components with known site of interaction			
α-Actinin-1,4	CT 1a/b, 5a/b	Enhances surface expression and signalling	64
Calmodulin	CT 5a/b	Regulates trafficking and signalling	64, 94
Ca ²⁺ channel (Ca _v 2.1)	CT 1a	Effector channel	64
CFTR-associated ligand	CT 1a	Regulates ubiquitination and trafficking	64
Filamin-A	CT 5a/b	ND	64
NECAB2	CT 5a/b	Reduces constitutive activity, increases agonist potency	64
GRK2	CT/IL2 1a/b	Reduces signalling	64
NHERF-2	CT 5a/b	Enhances G _q -mediated Ca ²⁺ responses, restricts coupling to G _q	64
NMDA receptors (NR1, NR2B)	CT 1a, 5a	Reciprocal inhibition	64
Norbin	CT 1a, 5a/b	Enhances surface expression and signalling	64
Optineurin	CT 1a, 5a	Reduces signalling	64
PP1γ	CT 1a, 5a/b	ND	64
Preso1	CT/Homer	Reduces signalling	95
Shank	CT/Homer	Couples to PSD and effector systems	63
Siah-1A	CT 1a, 5a/b	Mediates ubiquitination and trafficking	96
Tamalin	CT (PDZ-binding motif) 1a, 5a/b	Regulates surface expression and localization	64, 97
Calcineurin inhibitor	IL2 1a, 5a	Reduces endocytosis and signalling	64
Caveolin-1, -2β	IL1/IL3 1a	Reduces endocytosis, constitutive activity and signalling	64
Cytoshesin2‡	Tamalin	Regulates surface expression and localization	64, 97
IP3 receptor	Homer	Effector channel	63
PI3K enhancer	Homer	Coupling to PI3 kinase pathway	63
Ryanodine receptor	Homer	Effector channel	63
TRPC	Homer	Effector channel	63, 98
Peripheral components with unknown site of interaction			
Adenosine A ₂ receptor	5a/b	Synergistic signalling	64, 74, 77, 78
Ca ²⁺ -sensing receptor	1a	Promotes glutamate-induced receptor internalization	64
Dopamine D ₂ receptor	5a/b	ND	64, 74, 78
EphrinB2	1a, 5a/b	Synergistic signalling	64
μ-Opioid receptor	5a/b	Modulates endocytosis and pharmacology	76

The principal subunits mGlu₁ and mGlu₅ assemble into homodimers and possibly into mGlu₁–mGlu₅ heterodimers⁹. mGlu₁ and mGlu₅ subunit splice variants are denoted 1a, 5a and 5b. Homer proteins interact with the C terminus (CT) of mGlu₁ or mGlu₅ and exert both regulatory and scaffolding functions, similar as observed with the KCTD subunits of GABA_BRs. G proteins dock to a cavity between intracellular loop (IL) 2 and 3 of the principal subunits. Principal subunits, Homer and G-protein subunits constitute the core components of the receptor. As with GABA_BRs, it is expected that assembly with peripheral components generates a variety of molecularly distinct mGlu₁ and mGlu₅ complexes with unique functions and localization. In contrast to GABA_BRs, several reports suggest that mGluRs form heterodimers with class A GPCRs. To date, the peripheral components of mGluRs were mostly validated based on their effects on receptor signalling and trafficking. Few proteomic studies with native mGluRs were carried out and the interaction of most peripheral components with native receptors therefore still needs to be confirmed. A secondary interactor is indicated with ‡. ND, not determined.

appear to be the sole proteins that bind to both mGluRs and GABA_BRs. In contrast to the GABA_BRs, mGluRs are also reported to associate with other GPCRs, including for example, serotonin 5HT_{2A}, dopamine D₂, μ-opioid and adenosine receptors^{74–78}. Heteromeric complexes between mGluRs and other GPCRs enable a direct allosteric interaction between the two GPCR protomers. A notable example is the mGlu₂–5-HT_{2A} receptor⁷⁵. Association with mGlu₂ favours G_i over G_q signalling of the 5HT_{2A} protomer (that is, induces biased signalling, a concept that postulates that certain agonists or, in this case, association of 5HT_{2A} with mGlu₂ induces a distinct receptor conformation that selectively activates some but not all signalling pathways linked to the receptor⁷⁹). Hallucinogenic drugs therefore trigger unique cellular and behavioural responses through the 5-HT_{2A}–mGlu₂ complex. mGlu₂ agonists reverse this signalling bias and could potentially be used as antipsychotics to re-equilibrate G_i–G_q signalling of the complex⁷⁵. Functional interactions between striatal mGlu₅, adenosine A_{2A} and dopamine D₂ receptors suggest the possible usefulness of a combined use of A_{2A} and mGlu₅ antagonists in the treatment of Parkinson's disease and addiction^{74,77,78}. It is, however, not entirely clear whether these functional interactions represent direct heteromeric interactions⁷⁴ or signalling cross-talk.

Arguably the best-characterized mGluR interactome is that of group I mGluRs (Table 2). Cytoplasmic Homer proteins represent core building blocks of group I mGluRs and assume both regulatory and scaffolding functions. Homer-1a induces constitutive activity of mGluRs in response to increased neuronal activity, which scales down synaptic AMPA receptor expression⁸⁰. This homeostatic scaling mechanism keeps neuronal excitability constant even with increased neuronal activity⁸⁰. Genetic deletion of Homer-1a was recently shown to restore mGlu₅–Homer-1b scaffolds and to correct neocortical circuit dysfunctions and behaviours in a mouse model of fragile X syndrome⁸¹. In addition to these regulatory functions, Homer proteins also link the receptor to some peripheral building blocks (Table 2). Other peripheral building blocks directly interact with the C-terminal domain of principal subunits (Table 2). Several receptor-interacting proteins exhibit specificity for mGlu₁, mGlu₅ or variants thereof (Table 2). Heteromeric mGlu₁–mGlu₅ complexes may be physiologically relevant because they allow the simultaneous binding of mGlu₁- and mGlu₅-specific interacting proteins.

Targeting receptor protomers for drug discovery

Drug discovery approaches for class C GPCR have previously focused on the activation, inhibition and allosteric modulation of the core receptor

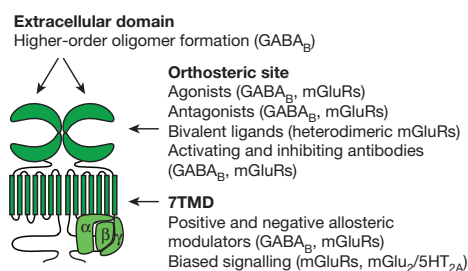
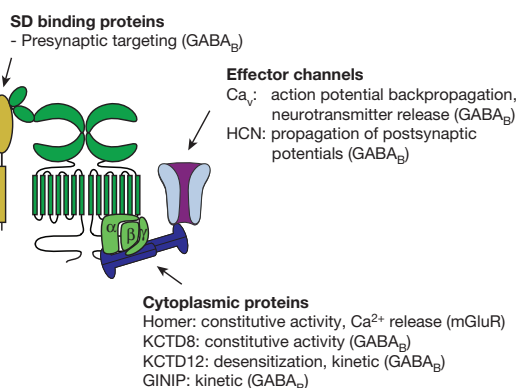
a Targeting principal subunits**b** Targeting associated proteins

Figure 4 | Opportunities for influencing mGluR and GABA_BR activity through orthosteric and allosteric binding sites or through interference with protein–protein interactions of receptor components.

a, Orthosteric and allosteric ligands acting at principal GABA_BR and mGluR subunits are available^{3–5}. To date, only the GABA_BR agonists baclofen and γ -hydroxy butyrate (GHB) are on the market³. Orthosteric ligands bind in the VFTDs, whereas allosteric modulators bind in the 7TMDs (modulators for GABA_BRs selectively act on GABA_{B2}). NAMs revealed at least two distinct lipophilic binding pockets in the mGlu 7TMD structures^{33,34}, which supports the existence of multiple allosteric sites in these receptors. Bivalent ligands that selectively bind to the mGlu₅– μ -opioid heteromer were shown to be antinociceptive in preclinical models⁷⁶. Auto-antibodies blocking the GABA binding site in GABA_{B1} have been identified⁸³. Allosteric modulation of the receptors

protomers^{3–5}. Accordingly, the available compounds and marketed drugs acting on mGluRs and GABA_BRs are small molecules that bind to orthosteric and allosteric sites in the VFTDs and 7TMDs (Fig. 4a). Notably, some allosteric modulators of mGlu₅ bias the signalling of orthosteric ligands⁸², which opens up novel therapeutic opportunities. The availability of X-ray structures of the orthosteric and allosteric binding pockets may allow broadening of the chemical and therapeutic spectrum of GABA_BR and mGluR ligands using structure-based drug design. As discussed above, the structures of the mGlu₁ (ref. 34) and mGlu₅ (ref. 33) 7TMDs in the presence of NAMs revealed topographically distinct highly lipophilic binding pockets of allosteric modulators. It will also be of importance to consider the structure of heterocomplexes when searching for new binding pockets and/or chemotypes. There is growing interest in using biologics and antibodies to prevent access of agonists to their binding sites or to stabilize receptor conformations and thereby control signalling. Camelid nanobodies or other antibody scaffolds that activate or modulate receptors by exploring the conformational landscape are particularly promising. The regulatory potential of antibodies is revealed by the identification of pathological auto-antibodies inhibiting the binding of GABA to GABA_BRs⁸³. Using antibodies, it may also be possible to specifically interfere with the formation of higher-order GABA_BR complexes (Fig. 4a). Moreover, bivalent ligands binding to both protomers or simultaneously to orthosteric and allosteric sites can be used to selectively target heterodimeric mGluRs, or to promote specific signalling pathways. For example, bivalent ligands that selectively interact with heteromeric mGlu₅– μ -opioid receptors exhibit potent antinociceptive effects⁷⁶ (Fig. 4a).

Targeting receptor-associated proteins for drug discovery

The delineation of multiprotein complexes with defined functions identifies additional receptor targets that may enable the design of drugs with more specific therapeutic effects and less side effects. Although it has proven difficult to pharmacologically interfere with protein–protein interactions using small molecules, brain- and cell-penetrable peptides have been shown to disrupt intracellular protein–protein interactions and to

by antibodies or ligands interfering with the formation of higher-order GABA_BR complexes have not been reported yet. **b**, No small compounds targeting protein–protein interactions in mGluRs or GABA_BRs are available. However, several proof-of-concept experiments reveal that a successful interference with associated proteins may be possible. For example, preventing the binding of proteins to the sushi domains of GABA_{B1a} interferes with presynaptic GABA_BR-mediated inhibition of glutamate release in acute hippocampal slice preparations⁸⁵. Preventing the association of HCN channels with GABA_BRs by deleting KCTD16 increases the inhibition of dopaminergic neurons in the VTA¹². Small peptides binding to the activity-dependent binding site of KCTD12 on G_{βγ} induce a pronounced desensitization of receptor-activated K⁺ currents, similar to KCTD12 (ref. 65).

be neuroprotective after intravenous injection⁸⁴. Preventing the binding of cytoplasmic Homer proteins to mGlu₁ and mGlu₅ could, for example, be used to reduce tonic receptor activity and to prevent Ca²⁺ release from intracellular stores (Fig. 4b). Preventing the binding of KCTD proteins to GABA_BRs would similarly provide a means to interfere with receptor desensitization, tonic receptor activity and coupling to effector channels (Fig. 4b). A drug that specifically impairs presynaptic GABA_BRs by interfering with the binding of proteins to the sushi domains would be desirable to selectively increase neurotransmitter release without indiscriminately antagonizing all GABA_BRs (Fig. 4b). As a proof of concept study, a recombinant sushi-domain protein was shown to interfere with presynaptic GABA_BR function in a dominant-negative manner in acute hippocampal slices from mice⁸⁵.

Outlook

There is still much to be discovered about macromolecular GABA_BR and mGluR complexes. Knowledge of the structure of both the VFTD and the 7TMD in various conformations already presents opportunities for tuning the activity of these receptors with small molecules or antibodies at various binding sites. However, it remains to be understood at the structural level how the VFTD controls the activity of the 7TMD. A detailed understanding will require resolution of the structure of the full-length dimer in its active and inactive state, with and without associated G protein. The physiological relevance of higher-order GABA_BR complexes is unknown and should be addressed once tools to disrupt these complexes are available. Proteomic studies, including reverse-affinity purifications of native receptor components, are needed to consolidate the receptor interactomes, in particular for mGluRs. Although initial insights into the components and their molecular interactions are now available, we still lack high-resolution structural information on the exact composition and stoichiometry of the complexes. Furthermore, we have very little understanding of the structural dynamics of these complexes as they perform their physiological functions. Most importantly, growing efforts are needed to understand where these complexes operate in neuronal networks and in cellular subdomains, which can be analysed by using protein complementation assays, chemical and photo-crosslinking techniques,

and proximity-induced ligation approaches⁸⁶. Given the rich interactome of class C GPCRs, future experiment should also address whether receptor complexes signal intracellularly and through G-protein-independent pathways, as observed for class A/B GPCRs⁵⁵. Understanding the functional properties of receptor complexes in their native environment will help to guide future efforts to define specific protein assemblies as targets for pharmacological intervention in disease.

Received 10 February; accepted 21 October 2016.

1. Gassmann, M. & Bettler, B. Regulation of neuronal GABA_B receptor functions by subunit composition. *Nat. Rev. Neurosci.* **13**, 380–394 (2012).
2. Niswender, C. M. & Conn, P. J. Metabotropic glutamate receptors: physiology, pharmacology, and disease. *Annu. Rev. Pharmacol. Toxicol.* **50**, 295–322 (2010).
3. Froestl, W. Chemistry and pharmacology of GABA_B receptor ligands. *Adv. Pharmacol.* **58**, 19–62 (2010).
4. Nickols, H. H. & Conn, P. J. Development of allosteric modulators of GPCRs for treatment of CNS disorders. *Neurobiol. Dis.* **61**, 55–71 (2014).
5. Bennett, K. A., Doré, A. S., Christopher, J. A., Weiss, D. R. & Marshall, F. H. Structures of mGluRs shed light on the challenges of drug development of allosteric modulators. *Curr. Opin. Pharmacol.* **20**, 1–7 (2015).
6. Kniazeff, J., Prézeau, L., Rondard, P., Pin, J. P. & Goudet, C. Dimers and beyond: the functional puzzles of class C GPCRs. *Pharmacol. Ther.* **130**, 9–25 (2011).
7. Marshall, F. H., Jones, K. A., Kaupmann, K. & Bettler, B. GABA_B receptors—the first 7TM heterodimers. *Trends Pharmacol. Sci.* **20**, 396–399 (1999).
8. Hannan, S., Gerrow, K., Triller, A. & Smart, T. G. Phospho-dependent accumulation of GABA_BRs at presynaptic terminals after NMDAR activation. *Cell Reports* **16**, 1962–1973 (2016).
9. Doumazane, E. *et al.* A new approach to analyze cell surface protein complexes reveals specific heterodimeric metabotropic glutamate receptors. *FASEB J.* **25**, 66–77 (2011).
10. Comps-Agrar, L. *et al.* The oligomeric state sets GABA_B receptor signalling efficacy. *EMBO J.* **30**, 2336–2349 (2011).
11. Schwenk, J. *et al.* Native GABA_B receptors are heteromultimers with a family of auxiliary subunits. *Nature* **465**, 231–235 (2010).
12. Schwenk, J. *et al.* Modular composition and dynamics of native GABA_B receptors identified by high-resolution proteomics. *Nat. Neurosci.* **19**, 233–242 (2016).
13. Galvez, T. *et al.* Allosteric interactions between GB1 and GB2 subunits are required for optimal GABA_B receptor function. *EMBO J.* **20**, 2152–2159 (2001).
14. Marsh, J. A. *et al.* Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* **153**, 461–470 (2013).
15. Doly, S. *et al.* GABA receptor cell-surface export is controlled by an endoplasmic reticulum gatekeeper. *Mol. Psychiatry* **21**, 480–490 (2016).
16. Burmakina, S., Geng, Y., Chen, Y. & Fan, Q. R. Heterodimeric coiled-coil interactions of human GABA_B receptor. *Proc. Natl Acad. Sci. USA* **111**, 6958–6963 (2014).
17. Maurel, D. *et al.* Cell-surface protein–protein interaction analysis with time-resolved FRET and snap-tag technologies: application to GPCR oligomerization. *Nat. Methods* **5**, 561–567 (2008).
18. Kniazeff, J. *et al.* Closed state of both binding domains of homodimeric mGlu receptors is required for full activity. *Nat. Struct. Mol. Biol.* **11**, 706–713 (2004).
19. Gavalas, A. *et al.* Segregation of family A G protein-coupled receptor protomers in the plasma membrane. *Mol. Pharmacol.* **84**, 346–352 (2013).
20. El Moustaine, D. *et al.* Distinct roles of metabotropic glutamate receptor dimerization in agonist activation and G-protein coupling. *Proc. Natl Acad. Sci. USA* **109**, 16342–16347 (2012).
21. Yin, S. *et al.* Selective actions of novel allosteric modulators reveal functional heteromers of metabotropic glutamate receptors in the CNS. *J. Neurosci.* **34**, 79–94 (2014).
22. Pandya, N. J. *et al.* Group 1 metabotropic glutamate receptors 1 and 5 form a protein complex in mouse hippocampus and cortex. *Proteomics* **16**, 2698–2705 (2016).
23. Comps-Agrar, L., Kniazeff, J., Brock, C., Trinquet, E. & Pin, J. P. Stability of GABA_B receptor oligomers revealed by dual TR-FRET and drug-induced cell surface targeting. *FASEB J.* **26**, 3430–3439 (2012).
24. Calebiro, D. *et al.* Single-molecule analysis of fluorescently labeled G-protein-coupled receptors reveals complexes with distinct dynamics and organization. *Proc. Natl Acad. Sci. USA* **110**, 743–748 (2013).
25. Sobolevsky, A. I., Rosconi, M. P. & Gouaux, E. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* **462**, 745–756 (2009).
26. Kunishima, N. *et al.* Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **407**, 971–977 (2000).
27. Tsuchiya, D., Kunishima, N., Kamiya, N., Jingami, H. & Morikawa, K. Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd³⁺. *Proc. Natl Acad. Sci. USA* **99**, 2660–2665 (2002).
28. Muto, T., Tsuchiya, D., Morikawa, K. & Jingami, H. Structures of the extracellular regions of the group II/III metabotropic glutamate receptors. *Proc. Natl Acad. Sci. USA* **104**, 3759–3764 (2007).
29. Geng, Y., Bush, M., Mosyak, L., Wang, F. & Fan, Q. R. Structural mechanism of ligand activation in human GABA_B receptor. *Nature* **504**, 254–259 (2013).
30. Kniazeff, J. *et al.* Locking the dimeric GABA_B G-protein-coupled receptor in its active state. *J. Neurosci.* **24**, 370–377 (2004).
31. Bessis, A.-S. *et al.* Closure of the Venus flytrap module of mGlu8 receptor and the activation process: insights from mutations converting antagonists into agonists. *Proc. Natl Acad. Sci. USA* **99**, 11097–11102 (2002).
32. Geng, Y. *et al.* Structure and functional interaction of the extracellular domain of human GABA_B receptor GBR2. *Nat. Neurosci.* **15**, 970–978 (2012).
33. Doré, A. S. *et al.* Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature* **511**, 557–562 (2014).
34. Wu, H. *et al.* Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science* **344**, 58–64 (2014).
35. Christopher, J. A. *et al.* Fragment and structure-based drug discovery for a class C GPCR: discovery of the mGlu5 negative allosteric modulator HTL14242 (3-chloro-5-[6-(5-fluoropyridin-2-yl)pyrimidin-4-yl]benzonitrile). *J. Med. Chem.* **58**, 6653–6664 (2015).
36. Chen, Y., Goudet, C., Pin, J.-P. & Conn, P. J. N-[4-chloro-2-[(1,3-dioxo-1,3-dihydro-2H-isoindol-2-yl)methyl]phenyl]-2-hydroxybenzamide (CPPHA) acts through a novel site as a positive allosteric modulator of group 1 metabotropic glutamate receptors. *Mol. Pharmacol.* **73**, 909–918 (2008).
37. Rovira, X. *et al.* Overlapping binding sites drive allosteric agonism and positive cooperativity in type 4 metabotropic glutamate receptors. *FASEB J.* **29**, 116–130 (2015).
38. Ango, F. *et al.* Agonist-independent activation of metabotropic glutamate receptors by the intracellular protein Homer. *Nature* **411**, 962–965 (2001).
39. Rodriguez, A. L. *et al.* A close structural analog of 2-methyl-6-(phenylethynyl)-pyridine acts as a neutral allosteric site ligand on metabotropic glutamate receptor subtype 5 and blocks the effects of multiple allosteric modulators. *Mol. Pharmacol.* **68**, 1793–1802 (2005).
40. Rook, J. M. *et al.* Biased mGlu5-positive allosteric modulators provide *in vivo* efficacy without potentiating mGlu5 modulation of NMDAR currents. *Neuron* **86**, 1029–1040 (2015).
41. Binet, V. *et al.* Common structural requirements for heptahelical domain function in class A and class C G protein-coupled receptors. *J. Biol. Chem.* **282**, 12154–12163 (2007).
42. Doumazane, E. *et al.* Illuminating the activation mechanisms and allosteric properties of metabotropic glutamate receptors. *Proc. Natl Acad. Sci. USA* **110**, E1416–E1425 (2013).
43. Olofsson, L. *et al.* Fine tuning of sub-millisecond conformational dynamics controls metabotropic glutamate receptors agonist efficacy. *Nat. Commun.* **5**, 5206 (2014).
44. Vafabakhsh, R., Levitz, J. & Isacoff, E. Y. Conformational dynamics of a class C G-protein-coupled receptor. *Nature* **524**, 497–501 (2015).
45. Huang, S. *et al.* Interdomain movements in metabotropic glutamate receptor activation. *Proc. Natl Acad. Sci. USA* **108**, 15480–15485 (2011).
46. Tateyama, M., Abe, H., Nakata, H., Saito, O. & Kubo, Y. Ligand-induced rearrangement of the dimeric metabotropic glutamate receptor 1α. *Nat. Struct. Mol. Biol.* **11**, 637–642 (2004).
47. Hlavackova, V. *et al.* Sequential inter- and intrasubunit rearrangements during activation of dimeric metabotropic glutamate receptor 1. *Sci. Signal.* **5**, ra59 (2012).
48. Hlavackova, V. *et al.* Evidence for a single heptahelical domain being turned on upon activation of a dimeric GPCR. *EMBO J.* **24**, 499–509 (2005).
49. Xue, L. *et al.* Major ligand-induced rearrangement of the heptahelical domain interface in a GPCR dimer. *Nat. Chem. Biol.* **11**, 134–140 (2015).
50. Maurice, P., Kamal, M. & Jockers, R. Asymmetry of GPCR oligomers supports their functional relevance. *Trends Pharmacol. Sci.* **32**, 514–520 (2011).
51. Monnier, C. *et al.* Trans-activation between 7TM domains: implication in heterodimeric GABA_B receptor activation. *EMBO J.* **30**, 32–42 (2011).

52. Goudet, C. *et al.* Asymmetric functioning of dimeric metabotropic glutamate receptors disclosed by positive allosteric modulators. *J. Biol. Chem.* **280**, 24380–24385 (2005).
53. Sun, B. *et al.* A negative allosteric modulator modulates GABA_B-receptor signalling through GB2 subunits. *Biochem. J.* **473**, 779–787 (2016).
54. Goudet, C. *et al.* Heptahelical domain of metabotropic glutamate receptor 5 behaves like rhodopsin-like receptors. *Proc. Natl Acad. Sci. USA* **101**, 378–383 (2004).
55. Lefkowitz, R. J. Arrestins come of age: a personal historical perspective. *Prog. Mol. Biol. Transl. Sci.* **118**, 3–18 (2013).
56. Iacovelli, L., Nicoletti, F. & De Biasi, A. Molecular mechanisms that desensitize metabotropic glutamate receptor signaling: an overview. *Neuropharmacology* **66**, 24–30 (2013).
57. Perroy, J., Adam, L., Qanbar, R., Chénier, S. & Bouvier, M. Phosphorylation-independent desensitization of GABA_B receptor by GRK4. *EMBO J.* **22**, 3816–3824 (2003).
58. Fourgeaud, L. *et al.* The metabotropic glutamate receptor mGluR5 is endocytosed by a clathrin-independent pathway. *J. Biol. Chem.* **278**, 12222–12230 (2003).
59. Fairfax, B. P. *et al.* Phosphorylation and chronic agonist treatment atypically modulate GABA_B receptor cell surface stability. *J. Biol. Chem.* **279**, 12565–12573 (2004).
60. Lujan, R. & Ciruela, F. GABA_B receptors-associated proteins: potential drug targets in neurological disorders? *Curr. Drug Targets* **13**, 129–144 (2012).
61. Bettler, B. & Tiao, J. Y. Molecular diversity, trafficking and subcellular localization of GABA_B receptors. *Pharmacol. Ther.* **110**, 533–543 (2006).
62. Benke, D. GABA_B receptor trafficking and interacting proteins: targets for the development of highly specific therapeutic strategies to treat neurological disorders? *Biochem. Pharmacol.* **86**, 1525–1530 (2013).
63. Bockaert, J., Perroy, J., Bécamel, C., Marin, P. & Fagni, L. GPCR interacting proteins (GIPs) in the nervous system: roles in physiology and pathologies. *Annu. Rev. Pharmacol. Toxicol.* **50**, 89–109 (2010).
64. Enz, R. Metabotropic glutamate receptors and interacting proteins: evolving drug targets. *Curr. Drug Targets* **13**, 145–156 (2012).
65. Turecek, R. *et al.* Auxiliary GABA_B receptor subunits uncouple G protein $\beta\gamma$ subunits from effector channels to induce desensitization. *Neuron* **82**, 1032–1044 (2014).
66. Tolkovsky, A. M. & Levitzki, A. Mode of coupling between the β -adrenergic receptor and adenylate cyclase in turkey erythrocytes. *Biochemistry* **17**, 3795 (1978).
67. Rajulu, M. *et al.* Pharmacological characterization of GABA_B receptor subtypes assembled with auxiliary KCTD subunits. *Neuropharmacology* **88**, 145–154 (2015).
68. Hanack, C. *et al.* GABA blocks pathological but not acute TRPV1 pain signals. *Cell* **160**, 759–770 (2015).
- Proteomic study reporting that GABA_{B1} interacts with TRPV1 channels in dorsal root ganglia and reverts the sensitized state of TRPV1 channels independent of G-protein signalling, possibly through allosteric interactions.**
69. Valdés, V. *et al.* Endoplasmic reticulum sorting and kinesin-1 command the targeting of axonal GABA_B receptors. *PLoS One* **7**, e44168 (2012).
70. Verhey, K. J. *et al.* Cargo of kinesin identified as JIP scaffolding proteins and associated signaling molecules. *J. Cell Biol.* **152**, 959–970 (2001).
71. Ludwig, A. *et al.* Calsynenins mediate TGN exit of APP in a kinesin-1-dependent manner. *Traffic* **10**, 572–589 (2009).
72. Laffray, S. *et al.* Impairment of GABA_B receptor dimer by endogenous 14-3-3 ζ in chronic pain conditions. *EMBO J.* **31**, 3239–3251 (2012).
73. Workman, E. R. *et al.* Rapid antidepressants stimulate the decoupling of GABA_B receptors from GIRK/Kir3 channels through increased protein stability of 14-3-3 η . *Mol. Psychiatry* **20**, 298–310 (2015).
74. Cabello, N. *et al.* Metabotropic glutamate type 5, dopamine D₂ and adenosine A_{2a} receptors form higher-order oligomers in living cells. *J. Neurochem.* **109**, 1497–1507 (2009).
75. Fribourg, M. *et al.* Decoding the signaling of a GPCR heteromeric complex reveals a unifying mechanism of action of antipsychotic drugs. *Cell* **147**, 1011–1023 (2011).
76. Akgün, E. *et al.* Ligands that interact with putative MOR–mGluR5 heteromer in mice with inflammatory pain produce potent antinociception. *Proc. Natl Acad. Sci. USA* **110**, 11595–11599 (2013).
77. Wright, S. R. *et al.* A critical role of striatal A2A R–mGlu5 R interactions in modulating the psychomotor and drug-seeking effects of methamphetamine. *Addict. Biol.* **21**, 811–825 (2016).
78. Beggiato, S. *et al.* Functional role of striatal A2A, D2, and mGlu5 receptor interactions in regulating striatopallidal GABA neuronal transmission. *J. Neurochem.* **138**, 254–264 (2016).
79. Kenakin, T. The effective application of biased signaling to new drug discovery. *Mol. Pharmacol.* **88**, 1055–1061 (2015).
80. Hu, J. H. *et al.* Homeostatic scaling requires group I mGluR activation mediated by Homer1a. *Neuron* **68**, 1128–1142 (2010).
81. Ronesi, J. A. *et al.* Disrupted Homer scaffolds mediate abnormal mGluR5 function in a mouse model of fragile X syndrome. *Nat. Neurosci.* **15**, 431–440 (2012).
82. Gentry, P. R., Sexton, P. M. & Christopoulos, A. Novel allosteric modulators of G protein-coupled receptors. *J. Biol. Chem.* **290**, 19478–19488 (2015).
83. Jain, A. & Balice-Gordon, R. Cellular, synaptic, and circuit effects of antibodies in autoimmune CNS synaptopathies. *Handb. Clin. Neurol.* **133**, 77–93 (2016).
84. Bach, A. *et al.* A high-affinity, dimeric inhibitor of PSD-95 bivalently interacts with PDZ1-2 and protects against ischemic brain damage. *Proc. Natl Acad. Sci. USA* **109**, 3317–3322 (2012).
85. Tiao, J. Y. *et al.* The sushi domains of secreted GABA_{B1} isoforms selectively impair GABA_B heteroreceptor function. *J. Biol. Chem.* **283**, 31005–31011 (2008).
86. Lowder, M. A., Appelbaum, J. S., Hobert, E. M. & Schepartz, A. Visualizing protein partnerships in living cells and organisms. *Curr. Opin. Chem. Biol.* **15**, 781–788 (2011).
87. Biermann, B. *et al.* The sushi domains of GABA_B receptors function as axonal targeting signals. *J. Neurosci.* **30**, 1385–1394 (2010).
88. Guetg, N. *et al.* NMDA receptor-dependent GABA_B receptor internalization via CaMKII phosphorylation of serine 867 in GABA_{B1}. *Proc. Natl Acad. Sci. USA* **107**, 13924–13929 (2010).
89. Kantamneni, S. *et al.* GISP: a novel brain-specific protein that promotes surface expression and function of GABA_B receptors. *J. Neurochem.* **100**, 1003–1017 (2007).
90. Vertkin, I. *et al.* GABA_B receptor deficiency causes failure of neuronal homeostasis in hippocampal networks. *Proc. Natl Acad. Sci. USA* **112**, E3291–E3299 (2015).
91. Gaillard, S. *et al.* GINIP, a G_{A1}-interacting protein, functions as a key modulator of peripheral GABA_B receptor-mediated analgesia. *Neuron* **84**, 123–136 (2014).
92. Xie, K. *et al.* G β 5 recruits R7 RGS proteins to GIRK channels to regulate the timing of neuronal inhibitory signaling. *Nat. Neurosci.* **13**, 661–663 (2010).
93. Tu, J. C. *et al.* Homer binds a novel proline-rich motif and links group 1 metabotropic glutamate receptors with IP3 receptors. *Neuron* **21**, 717–726 (1998).
94. Choi, K. Y., Chung, S. & Roche, K. W. Differential binding of calmodulin to group I metabotropic glutamate receptors regulates receptor trafficking and signaling. *J. Neurosci.* **31**, 5921–5930 (2011).
95. Hu, J. H. *et al.* Preso1 dynamically regulates group I metabotropic glutamate receptors. *Nat. Neurosci.* **15**, 836–844 (2012).
96. Ko, S. J. *et al.* PKC phosphorylation regulates mGluR5 trafficking by enhancing binding of Siah-1A. *J. Neurosci.* **32**, 16391–16401 (2012).
97. Kitano, J. *et al.* Tamalin, a PDZ domain-containing protein, links a protein complex formation of group 1 metabotropic glutamate receptors and the guanine nucleotide exchange factor cytohesins. *J. Neurosci.* **22**, 1280–1289 (2002).
98. Kim, S. J. *et al.* Activation of the TRPC1 cation channel by metabotropic glutamate receptor mGluR1. *Nature* **426**, 285–291 (2003).
99. Ferraguti, F. *et al.* Metabotropic glutamate receptor 8-expressing nerve terminals target subsets of GABAergic neurons in the hippocampus. *J. Neurosci.* **25**, 10520–10536 (2005).

Acknowledgements We thank M. Gassmann, J. Kniazeff and X. Rovira for discussions and help with the figures. This work was supported by grants of the Swiss Science Foundation (3100A0-117816) the National Center for Competences in Research (NCCR) ‘Synapsy, Synaptic Basis of Mental Health Disease’ to B.B., and by grants from the Agence National de la Recherche (ANR-12-BSV2-0015; ANR-13-RPIB-0009), the Fondation Recherche Médicale (FRM DEQ20130326522), the Fondation Bettencourt Schueller, and the Fond Unique Interministériel of the French government (FUI, Cell2Lead project) to J.-P.P.

Author Contributions J.-P.P. and B.B. wrote the manuscript together.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-P.P. (jppin@igf.cnrs.fr) or B.B. (bernhard.bettler@unibas.ch).

Reviewer Information Nature thanks K. Gregory, F. Marshall and the other anonymous reviewer(s) for their contribution to the peer review of this work.

The genomic basis of circadian and circalunar timing adaptations in a midge

Tobias S. Kaiser^{1,2,3†}, Birgit Poehn^{1,3}, David Szkiba², Marco Preussner⁴, Fritz J. Sedlazeck^{2†}, Alexander Zrim², Tobias Neumann^{1,2}, Lam-Tung Nguyen^{2,5}, Andrea J. Betancourt⁶, Thomas Hummel^{3,7}, Heiko Vogel⁸, Silke Dorner¹, Florian Heyd⁴, Arndt von Haeseler^{2,3,5} & Kristin Tessmar-Raible^{1,3}

Organisms use endogenous clocks to anticipate regular environmental cycles, such as days and tides. Natural variants resulting in differently timed behaviour or physiology, known as chronotypes in humans, have not been well characterized at the molecular level. We sequenced the genome of *Clunio marinus*, a marine midge whose reproduction is timed by circadian and circalunar clocks. Midges from different locations show strain-specific genetic timing adaptations. We examined genetic variation in five *C. marinus* strains from different locations and mapped quantitative trait loci for circalunar and circadian chronotypes. The region most strongly associated with circadian chronotypes generates strain-specific differences in the abundance of calcium/calmodulin-dependent kinase II.1 (CaMKII.1) splice variants. As equivalent variants were shown to alter CaMKII activity in *Drosophila melanogaster*, and *C. marinus* (*Cma*)-CaMKII.1 increases the transcriptional activity of the dimer of the circadian proteins *Cma*-CLOCK and *Cma*-CYCLE, we suggest that modulation of alternative splicing is a mechanism for natural adaptation in circadian timing.

Around the new or full moon, during a few specific hours surrounding low tide, millions of non-biting midges of the species *C. marinus* emerge from the sea to perform their nuptial dance. Adults live for only a few hours, during which they mate and oviposit. They must therefore emerge synchronously and—given that embryonic, larval and pupal development take place in the sea—at a time when the most extreme tides reliably expose the larval habitat. The lowest low tides occur predictably during specific days of the lunar month at a specific time of day. Consequently, adult emergence in *C. marinus* is under the control of circalunar and circadian clocks^{1,2}. Notably, although the lowest low tides recur invariably at a given location, their timing differs between geographic locations³. Consequently, *C. marinus* strains from different locations (Extended Data Fig. 1a) show local adaptation in circadian and circalunar emergence times (Extended Data Fig. 1b, c). Crosses between the Jean and Por strains showed that the differences in circadian and circalunar timing are genetically determined^{4,5} and largely explained by two circadian and two circalunar quantitative trait loci (QTLs)⁶.

Studies on timing variation or chronotypes in animals and humans have often focused on candidate genes from the circadian transcription-translational oscillator. In *D. melanogaster*, polymorphisms in the core circadian clock genes *period*, *timeless* and *cryptochrome* are associated with adaptive differences in temperature compensation⁷, photo-responsiveness of the circadian clock⁸ and emergence rhythms⁹. While these studies offer insights into the evolution of known circadian-clock molecules, genome-wide association studies^{10,11} and other forward genetic approaches (reviewed in ref. 12) are essential to provide a comprehensive, unbiased assessment of natural timing variation, for instance underlying human sleep-phase disorders. While the adaptive nature of human chronotypes remains unclear, the chronotypes of *C. marinus* represent evolutionary adaptations to their habitat.

Our study aimed to identify the genetic basis of *C. marinus* adaptation to its specific ecological 'timing niche'. In addition, the genetic dissection of adaptive natural variants of non-circadian rhythms¹³, as also present in *C. marinus*, may provide an entry point into their unknown molecular mechanisms.

As a starting point for these analyses, we sequenced, assembled, mapped and annotated a *C. marinus* reference genome.

The *Clunio* genome and QTLs for timing

Our reference genome CLUMA_1.0 of the Jean laboratory strain contained 85.6 Mb of sequence (Table 1), close to the previous flow-cytometry-based estimate of 95 Mb⁶, underlining that chironomids generally have small genomes^{14–16}. The final assembly has a scaffold N50 of 1.9 Mb. Genome-wide genotyping of a mapping family with restriction-site associated DNA sequencing allowed 92% of the reference sequence to be consistently anchored along a genetic linkage map (Fig. 1a and Extended Data Fig. 2), improving the original linkage map (Supplementary Methods 5). Automated genome annotation resulted in 21,672 gene models. Protein similarity and available transcripts support 14,041 gene models (Supplementary Table 1), within the range of gene counts for *D. melanogaster* (15,507) and *Anopheles gambiae* (13,460). Thus, the very small *C. marinus* genome appears to be complete (Table 1, Extended Data Fig. 3, Supplementary Note 1 and Supplementary Table 2). The *C. marinus* reference genome makes chironomids the third dipteran subfamily with an annotated genome reconstructed to chromosome scale (Fig. 1a and Extended Data Figs 2, 3b–f).

We performed a basic genome characterization and comparison to other dipterans. We delineated the five *C. marinus* chromosome arms (Supplementary Note 2, Extended Data Fig. 3c and Supplementary Table 3) and homologized them to *D. melanogaster* and *A. gambiae* by synteny comparisons (Extended Data Figs 3 and 4, Supplementary Note 2 and

¹Max F. Perutz Laboratories, University of Vienna, Campus Vienna Biocenter, Dr. Bohr-Gasse 9/4, A-1030 Vienna, Austria. ²Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna and Medical University of Vienna, Dr. Bohr-Gasse 9, A-1030 Vienna, Austria. ³Research Platform 'Rhythms of Life', University of Vienna, A-1030 Vienna, Austria. ⁴Department of Biology, Chemistry, Pharmacy, Institute of Chemistry and Biochemistry, FU Berlin, D-14195 Berlin, Germany. ⁵Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, A-1030 Vienna, Austria. ⁶Institute of Population Genetics, Department of Biomedical Sciences, University of Veterinary Medicine Vienna, Josef-Baumann-Gasse 1, A-1210 Vienna, Austria. ⁷Department of Neurobiology, Faculty of Life Sciences, University of Vienna, A-1090 Vienna, Austria. ⁸Department of Entomology, Max Planck Institute for Chemical Ecology, Hans-Knöll-Straße 8, D-07745 Jena, Germany. ⁹Present addresses: Max Planck Institute for Evolutionary Biology, August-Thienemann-Straße 2, D-24306 Plön, Germany (T.S.K.); Department of Computer Science, Johns Hopkins University, Baltimore, Maryland 21211, USA (F.J.S.).

Table 1 | Comparison of the *C. marinus* genome assembly with published model insect genomes

	<i>Clunio marinus</i>	<i>Danaus plexippus</i> ⁴³	<i>Tribolium castaneum</i> ⁴⁴	<i>Apis mellifera</i> ⁴⁵
Total bases (Mb)	86	278	160	236
Mapped sequence (%)	92	NA	90	79
Scaffold N50 (Mb)	1.9	0.2	1	0.4
Contig N50 (kb)	79	50	41	41
AT content (%)	68.3	68.4	67	66
Completeness (%)	98.0	98.5	NA	NA
Platform	Illumina	Illumina + 454	Sanger + BAC	Sanger + BAC

Machine readable super-scaffolding data and the computer source code for the removal of repeated edges are supplied as source data files. 454, 454 sequencing (Life Sciences); BAC, bacterial artificial chromosome.

Supplementary Table 3). We also found the ZW-like sex-linked locus in *C. marinus*⁶ outside the X chromosome homologue (Supplementary Note 2) and detected an elevated rate of chromosomal re-arrangement (Fig. 1a, Supplementary Note 3 and Extended Data Figs 2, 3b–f, 4). Taken together, the *C. marinus* reference genome appears well assembled.

As the next step towards identifying the molecular basis of circadian and circalunar timing adaptations in *C. marinus*, we refined the previously identified timing QTL positions⁶ based on the new high-density, restriction-site-associated DNA sequencing markers (Supplementary Table 4 and Supplementary Note 4) and determined the reference sequence corresponding to the QTL confidence intervals (Fig. 1, orange and cyan bars; and Supplementary Table 4). None of the core circadian clock genes were found to be located within these QTLs (Fig. 1a). Only *timeout/timeless2*, a *timeless* homologue with a minor role in circadian clock resetting¹⁷, is located within the QTLs.

Genetic variation in *Clunio* timing strains

We then re-sequenced the Por and Jean strains (Extended Data Fig. 1), for which the initial QTL analysis was performed⁶. Two pools of 300 field-caught individuals were sequenced at >240× coverage (Supplementary Table 5). Mapping reads against the reference genome identified 1,010,052 single nucleotide polymorphisms (SNPs), 72% of which were present in both the Por and Jean strains. Based on all SNPs we determined genetic differentiation (F_{ST}), genetic diversity (θ) and short-range linkage disequilibrium (measured as r^2) (Fig. 1b and Extended Data Figs 3c, 5a, b).

Genome-wide genetic differentiation between the Por and Jean strains is moderate ($F_{ST}=0.11$), providing a good basis for screening the genome for local timing adaptation based on genetic divergence. According to QTL analysis, the two circadian QTLs explain 85% of the daily timing difference and the two circalunar QTLs explain the entire monthly timing difference (Supplementary Table 4 and ref. 6). As each locus therefore has a strong effect on timing, selection against maladapted alleles must be strong and timing loci should be strongly differentiated.

Within the confidence intervals of the QTLs, 158 SNPs and 106 indels (insertions or deletions) are strongly differentiated ($F_{ST} \geq 0.8$; Fig. 1b and Extended Data Fig. 5; SNPs, red dots in F_{ST} panels, for genome-wide comparison see Supplementary Note 5). We compiled a list of candidate genes for circadian and circalunar timing adaptations based on their proximity to differentiated SNPs and indels in the QTLs (Supplementary Table 6). The candidate genes neither comprise core circadian clock genes (*timeless2/timeout*, max. $F_{ST} \leq 0.5$; average $F_{ST}=0.07$), nor are enriched for any particular pathway (gene ontology-term analysis; Supplementary Table 7).

Timing phenotype with genotype correlation

Assuming that the alleles associated with timing adaptation probably originated from standing genetic variation (Supplementary Note 5),

genetic variation at timing loci should not vary freely between strains, but rather strains with similar timing should share functionally relevant alleles. To identify such loci, we extended the genomic screen to three additional strains: from Vigo (Vigo), Helgoland (He) and Bergen (Ber; Extended Data Fig. 1 and Supplementary Tables 5, 8). We then tested all five sequenced strains for correlations between genetic differentiation (F_{ST}) and timing differences, or geographic distances as a null model (Supplementary Table 8).

Overall, genome-wide genetic differentiation was not correlated with circadian ($r=0.10$, $P=0.31$) or circalunar ($r=0.56$, $P=0.12$) timing differences, but with geographic distance ('isolation by distance'; $r=0.88$, $P=0.008$). Against this genomic background signal of isolation by distance, we screened the genome in 5-kb sliding windows for peaks of correlation between genetic differentiation and timing, resulting in a correlation score (Fig. 1b and Extended Data Fig. 5a, b, CS panels, score ranging from 0 to 5; for details see Methods). Combining the evidence from the Por versus Jean strain F_{ST} screen (Supplementary Table 6) with these patterns of correlation between timing and genetic divergence reduced the candidate gene list to 49 genes (Supplementary Table 9).

Of particular note, a single region in circadian QTL C2 was strikingly differentiated (Fig. 1b). In this region, linkage disequilibrium in the Por strain was significantly elevated (permutation test; $P=0.002$), and genetic diversity significantly decreased in some stretches (permutation test; $P=0.037$ and 0.020), compared to the Por genome average. This may indicate a recent episode of selection in Por, potentially during timing adaptation, as this region is also strongly enriched for timing-correlated polymorphisms (Fig. 1b, CS panel). The most extreme values of genetic differentiation, genetic diversity and timing correlation localize to the *CaMKII.1* locus and the anterior section of a gene homologous to the *big bang* (*bbg*) gene.

CaMKII affects the circadian core clock

Not only does the *CaMKII.1* locus harbour the highest number of differentiated polymorphisms (Supplementary Table 9), but CaMKII has also been shown to affect circadian timing. Mouse CaMKII α phosphorylates CLOCK and facilitates its dimerization with BMAL *in vivo*¹⁸. Mice with inactive, kinase-dead CaMKII α ^{K42R} have dampened circadian rhythms and a lengthened circadian free-running period¹⁸. CaMKII also phosphorylates the CLOCK protein¹⁹ in the *D. melanogaster* S2 cell line, and *in vivo* inhibition of *Dme*-CaMKII in a sensitized background with reduced Ca^{2+} levels lengthens the circadian free-running period²⁰, suggesting that the role of CaMKII in circadian timing is conserved across animals.

To determine whether CaMKII can also affect the circadian core clock in *C. marinus*, we tested the effect of *Cma*-CaMKII.1 in a cell-based assay using *D. melanogaster* S2 cells^{19,21}. We repeated previous experiments¹⁹ showing that the chemical inhibition of endogenous *Dme*-CaMKII reduces the amount of generated luciferase (Extended Data Fig. 6a), whereas addition of a $[Ca^{2+}]$ -independent, and therefore constitutively active, variant of CaMKII (mouse, T286D) increases luciferase amounts (Extended Data Fig. 6b). Then we generated constructs for *C. marinus* clock, *C. marinus* cycle, and mutated kinase-dead (K42R) and $[Ca^{2+}]$ -independent (T286D) versions of *Cma*-CaMKII.1. Transfection of *Cma*-clock and *Cma*-cycle into *D. melanogaster* S2 cells leads to luciferase activity driven from the 3X69 promoter derived from the *Dme*-period promoter (Fig. 2a). The addition of $[Ca^{2+}]$ -independent *Cma*-CaMKII.1^{T286D} leads to a substantial increase in the luciferase signal (Fig. 2a), whereas addition of the kinase-dead *Cma*-CaMKII.1^{K42R} does not enhance luciferase activity (Fig. 2a). These data suggest that CaMKII kinase activity enhances E-box dependent transcription, as indicated by luciferase production driven by the 3X69promoter, via the CLOCK–CYCLE dimer in *C. marinus*.

CaMKII.1 splicing correlates with timing

We then investigated how polymorphisms in the *Cma*-CaMKII.1 locus affect the enzyme. We found two *CaMKII.1* alleles: one in the early

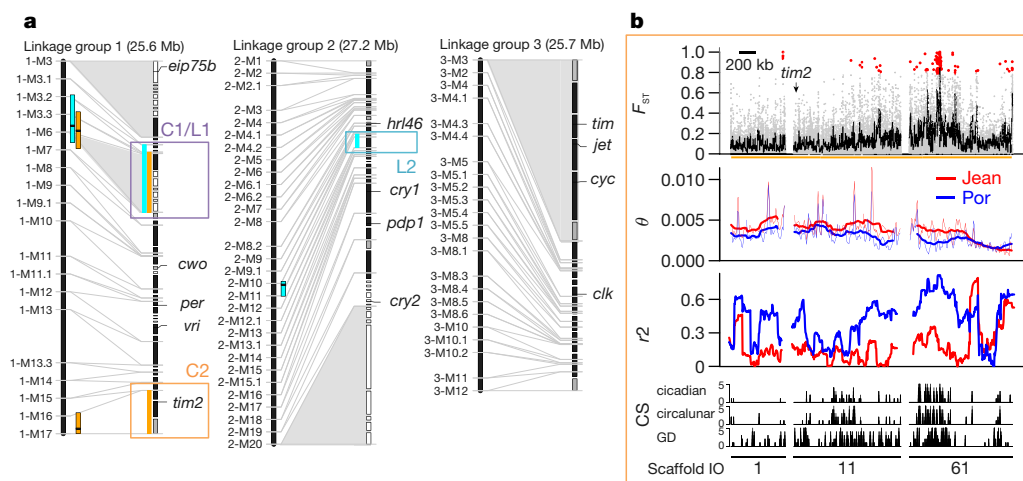


Figure 1 | Identification of candidate regions in the timing QTLs by combined genetic and molecular maps. **a**, The three linkage groups of *C. marinus* with reference scaffolds (right) anchored on a genetic linkage map (left). Scaffolds which are ordered and oriented, black bars; not oriented, grey bars; neither ordered nor oriented, white bars. Grey shadings, large non-recombining regions. QTLs, circadian (orange), circalunar (cyan). One circadian and circalunar QTL overlap, resulting in three physical QTL regions (C1/L1, C2 and L2, in purple, orange and cyan, respectively). **b**, Population genomic analysis of QTL C2. Analysis of

Por and Jean strains (in blue and red, respectively, in middle two panels). Top panel, genetic differentiation for single SNPs (red dots) and in 5-kb windows (black line). Second panel, genetic diversity (θ) in 20-kb (thin line) and 200-kb (thick line) windows. Third panel, linkage disequilibrium (r^2) in 100-kb windows. Bottom panel, correlation score (CS) for genetic differentiation with values for circadian timing (top), circalunar timing (middle) and geographic distance (bottom) for Vigo, Jean, Por, He and Ber strains. Bottom numbers, scaffold IDs. For further details, including QTLs C1/L1 and L2, see Extended Data Fig. 5a, b.

emerging Por, He and Ber strains, and another in the late emerging Jean and Vigo strains. Most strain-specific polymorphisms are located in introns (Fig. 2b, c and Supplementary Table 9). If these polymorphisms were meaningful, then they should affect *CaMKII.1* expression and/or splicing. *Cma-CaMKII.1* has four functional domains²² (Fig. 2b). The majority of differentiated polymorphisms cluster in the region of the variable linker domain (Fig. 2b, c), including a 125-bp insertion (red dot in Fig. 2c; Extended Data Fig. 7). We identified four alternatively spliced full-length transcripts of *Cma-CaMKII.1* (RA–RD), which differ in the length of the linker (Fig. 2b). High-coverage RNA sequencing gave evidence for differential exon usage between the Jean and Por strains, as well as for previously non-annotated exons within the variable linker region (Extended Data Fig. 6c). PCR and Sanger sequencing confirmed several partial transcripts of additional splice variants of the linker region (RE–RO; Fig. 2b). We used transcript-specific qPCR to quantify all transcripts from third instar larvae. Generally, transcripts RE–RO are expressed at very low levels. Of those, only RO showed

quantifiable expression differences between the Jean and Por strains (Fig. 3a and Extended Data Fig. 6d). Importantly, transcript-specific qPCR confirmed significant differential expression of the major transcripts in the Jean versus Por strains (Fig. 3a, Extended Data Fig. 6d), matching the RNA-sequencing (RNA-seq) data (Extended Data Fig. 6c). Consistently, variants with long linkers (RA, RB) showed higher expression in the Por strain, whereas shorter variants (RD, RO) showed higher expression in the Jean strain (Fig. 3a and Extended Data Fig. 6c, d).

If the detected differences in the abundance of *CaMKII.1* splice variants are associated with the timing differences, they should be directly caused by the strain-specific polymorphisms at the *CaMKII.1* locus. In order to test this, we generated minigenes that contained the alternatively spliced linker region of the *CaMKII.1* locus from either the Jean or Por strains. The two minigenes were transfected into cells of the *D. melanogaster* S2R+ cell line and expression of splice variants was analysed by radioactive RT–PCR (Fig. 3b, c). We detected four variants, corresponding to splice variants RB, RC, RD and RO. All variants

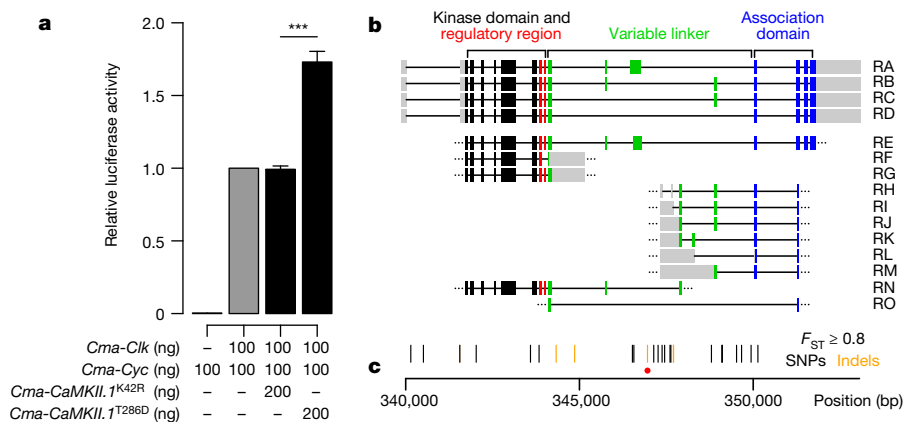


Figure 2 | *CaMKII.1* regulates *Clk* and *Cyc* transcriptional activity and exhibits strain specific splice variants. **a**, Additional *C. marinus* *CaMKII.1* increases the transcriptional activity of *C. marinus* *Clk* and *Cyc* in a *D. melanogaster* S2 cell luciferase assay using the 3X69 E-box containing enhancer (*period* 3X69–*luc* (ref. 21)). Data are represented as mean \pm s.e.m.; two-sided Welch two-sample *t*-test; biological replicates,

$n = 5$, except for no *clk* control, $n = 3$, each biological replicate represents the average of three preparation replicates. *** $P < 0.0005$. **b**, Exons of full (RA–RD) and partial (RE–RO) *Cma-CaMKII.1* transcripts. **c**, Distribution of SNPs (black), indels (orange) and a 125-bp insertion (red dot) along the *Cma-CaMKII.1* locus, all with $F_{ST} \geq 0.8$.

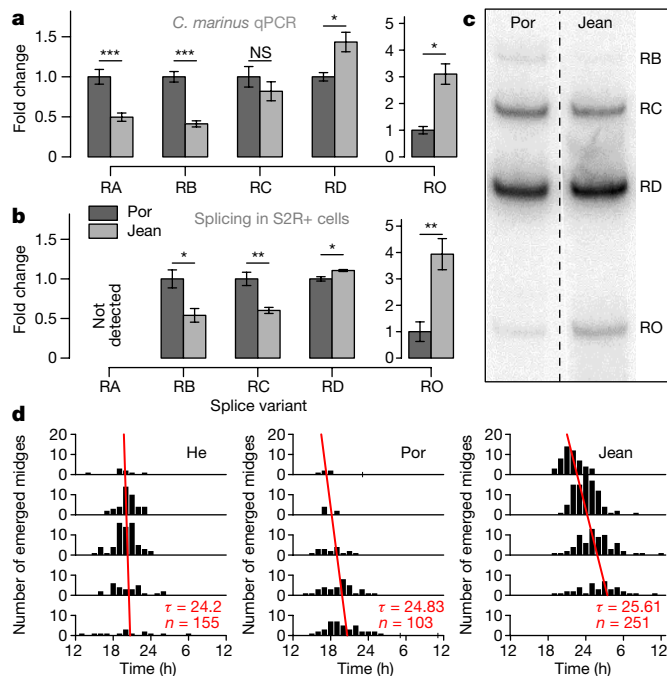


Figure 3 | Differential *CaMKII.1* splicing depends on sequence differences in the *CaMKII.1* locus and correlates with endogenous circadian period lengths. **a**, qPCR values for *CaMKII.1* splice variants from Por and Jean strains, normalized to Por (for non-normalized data, see Extended Data Fig. 6d). Data are represented as mean \pm s.e.m.; Por, $n = 9$ biological replicates; Jean, $n = 10$; RO, Por, $n = 3$; Jean $n = 8$; RO was not detected in six Por biological replicates, suggesting an even larger expression difference; two-sided Wilcoxon rank-sum test; * $P < 0.05$; ** $P < 0.005$; *** $P < 0.0005$; NS, not significant; Holm correction for multiple testing. For RNA-seq data quantification see Extended Data Fig. 6c. **b**, Differential splicing of the *CaMKII.1* linker region in *D. melanogaster* S2R+ cells, normalized to Por, $n = 7$ biological replicates; two-sided two-sample *t*-test, otherwise as **a**. **c**, Representative phosphorimaging gel sections as quantified for **b**, two separate lanes from the same gel (for full gel, see Source Data). **d**, Free-running rhythm of adult emergence under constant dim white light (approximately 100 lx). He and Por share *CaMKII.1* alleles, while Jean has the other allele. To calculate the free-running period, time between subsequent emergence peaks was averaged, weighting each peak by the number of individuals.

showed the same strain-specific abundance differences in the S2R+ cell assay as in the *C. marinus* strains *in vivo* (Fig. 3a, b). As the cellular context is the same for both the Jean and Por minigenes in the S2R+ assay, *trans*-acting elements can be excluded as the cause of differential splicing, implying that it is a direct result of the genomic sequence differences at the *Cma-CaMKII.1* locus. While splice variants RB, RC and RD and their constituting exons are conserved in *D. melanogaster* (see Flybase annotations and ref. 23), a *D. melanogaster* RA counterpart does not exist. This may explain why this variant is undetectable in S2R+ cells (Fig. 3b).

From splice variants to timing differences

CaMKII linker-length variants have been investigated in several species. *D. melanogaster* *CaMKII* isoforms corresponding to the RB, RC and RD variants of *C. marinus* have different substrate affinities and rates of target phosphorylation²³. These activity differences are explained by the fact that *CaMKII* functions as a dodecamer, and the linker length determines the compactness and thus the substrate accessibility of the holoenzyme—enzymes with long linkers have higher activity. This structure–function relationship is possibly universal, as it is conserved between humans and *C. elegans*^{22,24}.

Inactivation or inhibition of *CaMKII* lengthens circadian periods in mouse and fruitflies^{18,20}. A connection between circadian period

length and phase of activity in light–dark cycles is known from mutations in *period* in *D. melanogaster*²⁵ and human chronotypes²⁶. These findings imply that in *C. marinus* the more active and more readily Ca^{2+} -activated, long-linker *CaMKII.1* variants should advance adult emergence by shortening the circadian clock period. Indeed, we find that the early emerging Por and He strains, which possess the same long-linker biased *CaMKII.1* alleles, have shorter free-running circadian clock periods than the late emerging Jean strain (Fig. 3d).

Integrating our results with those from the aforementioned literature, we propose that regulation of the ratio of *CaMKII.1* splice variants constitutes an evolutionary mechanism to adapt circadian timing (Extended Data Fig. 8): differences in the genomic sequence of *CaMKII.1* lead to differential *CaMKII.1* splicing and activity. Among a number of possible targets, this influences CLOCK–CYCLE dimer-dependent transcription, which in turn affects circadian period length and ultimately results in differences in adult emergence time.

Discussion

Annual, lunar and tidal rhythms, as well as natural timing variation between individuals, are important and widespread phenomena that are poorly understood. The *C. marinus* reference genome and the genetic variation panel for five strains with differing circadian and circalunar timing establish new resources for further studies of these topics.

We identified *C. marinus* orthologues for all core circadian clock genes, none of which appear to be involved in circadian or circalunar timing adaptations. For circalunar timing, this supports the molecular independence of the circalunar clock from the circadian clock, as reported for *Platynereis dumerilii*²⁷.

For circadian timing, strain-specific modulation in alternative splicing of *CaMKII.1* emerges as a possible mechanism for natural adaptation. In the light of previous experiments in *D. melanogaster* and mice^{18–20,23}, it seems most likely that differences in *CaMKII* activity of the different splice forms lead to circadian timing differences via phosphorylation of CLOCK–CYCLE (Extended Data Fig. 8).

It is also conceivable that *CaMKII* affects circadian timing via other targets. For example, *CaMKII* is known to phosphorylate the cAMP response element binding protein (CREB)^{28,29}. CREB is linked to the circadian clock by cAMP response elements (CRE) in the promoters of the *period* and *timeless* genes^{30,31}, and by physical interaction of the CREB-binding protein (CBP) with CREB, CLOCK and CYCLE^{32,33}. Furthermore, one of the most well-studied roles of *CaMKII* is the morphological modulation of neuronal plasticity and connectivity^{34–36}. Such changes in connectivity have been increasingly implicated as part of the circadian timing mechanism in *D. melanogaster* and mammals³⁷. Interestingly, the role of *CaMKII* in shaping neuronal connectivity has also been suggested to link to several neuropsychiatric diseases³⁸, which often co-occur with chronobiological disruptions^{39–42}. Further studies are needed to determine whether the modulation of *CaMKII* activity constitutes a molecular link between these phenomena.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 March 2015; accepted 10 October 2016.

Published online 21 November 2016.

- Neumann, D. Die lunare und tägliche Schlüpfperiodik der Mücke *Clunio*. Steuerung und Abstimmung auf die Gezeitenperiodik. *Z. Vgl. Physiol.* **53**, 1–61 (1966).
- Neumann, D. Temperature compensation of circasemilunar timing in the intertidal insect *Clunio*. *J. Comp. Physiol.* **163**, 671–676 (1988).
- ADMIRALTY Tide Tables (UKHO, 2014).
- Neumann, D. Genetic adaptation in emergence time of *Clunio* populations to different tidal conditions. *Helgoländer Wiss. Meeresunters.* **15**, 163–171 (1967).
- Kaiser, T. S., Neumann, D. & Heckel, D. G. Timing the tides: genetic control of diurnal and lunar emergence times is correlated in the marine midge *Clunio marinus*. *BMC Genet.* **12**, 49 (2011).

6. Kaiser, T. S. & Heckel, D. G. Genetic architecture of local adaptation in lunar and diurnal emergence times of the marine midge *Clunio marinus* (Chironomidae, Diptera). *PLoS One* **7**, e32092 (2012).
7. Sawyer, L. A. *et al.* Natural variation in a *Drosophila* clock gene and temperature compensation. *Science* **278**, 2117–2120 (1997).
8. Sandrelli, F. *et al.* A molecular basis for natural selection at the *timeless* locus in *Drosophila melanogaster*. *Science* **316**, 1898–1900 (2007).
9. Pegoraro, M. *et al.* Molecular evolution of a pervasive natural amino-acid substitution in *Drosophila cryptochrome*. *PLoS One* **9**, e86483 (2014).
10. Lane, J. M. *et al.* Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. *Nat. Commun.* **7**, 10889 (2016).
11. Hu, Y. *et al.* GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Commun.* **7**, 10448 (2016).
12. Jones, C. R., Huang, A. L., Ptáček, L. J. & Fu, Y.-H. Genetic basis of human circadian rhythm disorders. *Exp. Neurol.* **243**, 28–33 (2013).
13. Tessmar-Raible, K., Raible, F. & Arboleda, E. Another place, another timer: marine species and the rhythms of life. *BioEssays* **33**, 165–172 (2011).
14. Gusev, O. *et al.* Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nat. Commun.* **5**, 4784 (2014).
15. Cornette, R. *et al.* Chironomid midges (Diptera, chironomidae) show extremely small genome sizes. *Zoolog. Sci.* **32**, 248–254 (2015).
16. Kelley, J. L. *et al.* Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. *Nat. Commun.* **5**, 4611 (2014).
17. Benna, C. *et al.* *Drosophila timeless2* is required for chromosome stability and circadian photoreception. *Curr. Biol.* **20**, 346–352 (2010).
18. Kon, N. *et al.* CaMKII is essential for the cellular clock and coupling between morning and evening behavioral rhythms. *Genes Dev.* **28**, 1101–1110 (2014).
19. Weber, F., Hung, H. C., Maurer, C. & Kay, S. A. Second messenger and Ras/MAPK signalling pathways regulate CLOCK/CYCLE-dependent transcription. *J. Neurochem.* **98**, 248–257 (2006).
20. Harrisingh, M. C., Wu, Y., Lnenicka, G. A. & Nitabach, M. N. Intracellular Ca²⁺ regulates free-running circadian clock oscillation *in vivo*. *J. Neurosci.* **27**, 12489–12499 (2007).
21. Nawathean, P. & Rosbash, M. The doubletime and CKII kinases collaborate to potentiate *Drosophila* PER transcriptional repressor activity. *Mol. Cell* **13**, 213–223 (2004).
22. Chao, L. H. *et al.* A mechanism for tunable autoinhibition in the structure of a human Ca²⁺/calmodulin-dependent kinase II holoenzyme. *Cell* **146**, 732–745 (2011).
23. GuptaRoy, B. *et al.* Alternative splicing of *Drosophila* calcium/calmodulin-dependent protein kinase II regulates substrate specificity and activation. *Brain Res. Mol. Brain Res.* **80**, 26–34 (2000).
24. Chao, L. H. *et al.* Intersubunit capture of regulatory segments is a component of cooperative CaMKII activation. *Nat. Struct. Mol. Biol.* **17**, 264–272 (2010).
25. Hamblen-Coyle, M. J., Wheeler, D. A., Rutilla, J. E., Rosbash, M. & Hall, J. C. Behavior of period-altered circadian rhythm mutants of *Drosophila* in light: dark cycles (Diptera: Drosophilidae). *J. Insect Behav.* **5**, 417–446 (1992).
26. Brown, S. A. *et al.* Molecular insights into human daily behavior. *Proc. Natl Acad. Sci. USA* **105**, 1602–1607 (2008).
27. Zantke, J. *et al.* Circadian and circalunar clock interactions in a marine annelid. *Cell Rep.* **5**, 99–113 (2013).
28. Sun, P., Enslen, H., Myung, P. S. & Maurer, R. A. Differential activation of CREB by Ca²⁺/calmodulin-dependent protein kinases type II and type IV involves phosphorylation of a site that negatively regulates activity. *Genes Dev.* **8**, 2527–2539 (1994).
29. Wu, X. & McMurray, C. T. Calmodulin kinase II attenuation of gene transcription by preventing cAMP response element-binding protein (CREB) dimerization and binding of the CREB-binding protein. *J. Biol. Chem.* **276**, 1735–1741 (2001).
30. Belvin, M. P., Zhou, H. & Yin, J. C. P. The *Drosophila* dCREB2 gene affects the circadian clock. *Neuron* **22**, 777–787 (1999).
31. Okada, T. *et al.* Promoter analysis for daily expression of *Drosophila* timeless gene. *Biochem. Biophys. Res. Commun.* **283**, 577–582 (2001).
32. Lim, C. *et al.* Functional role of CREB-binding protein in the circadian clock system of *Drosophila melanogaster*. *Mol. Cell Biol.* **27**, 4876–4890 (2007).
33. Lee, Y. *et al.* Coactivation of the CLOCK-BMAL1 complex by CBP mediates resetting of the circadian clock. *J. Cell Sci.* **123**, 3547–3557 (2010).
34. Kalil, K., Li, L. & Hutchins, B. I. Signaling mechanisms in cortical axon growth, guidance, and branching. *Front. Neuroanat.* **5**, 62 (2011).
35. Hell, J. W. CaMKII: claiming center stage in postsynaptic function and organization. *Neuron* **81**, 249–265 (2014).
36. McVicker, D. P., Millette, M. M. & Dent, E. W. Signaling to the microtubule cytoskeleton: an unconventional role for CaMKII. *Dev. Neurobiol.* **75**, 423–434 (2015).
37. Bosler, O., Girardet, C., Franc, J.-L., Becquet, D. & François-Bellan, A.-M. Structural plasticity of the circadian timing system. An overview from flies to mammals. *Front. Neuroendocrinol.* **38**, 50–64 (2015).
38. Robison, A. J. Emerging role of CaMKII in neuropsychiatric disease. *Trends Neurosci.* **37**, 653–662 (2014).
39. Wulff, K., Gatti, S., Wettstein, J. G. & Foster, R. G. Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nat. Rev. Neurosci.* **11**, 589–599 (2010).
40. Levandovski, R. *et al.* Depression scores associate with chronotype and social jetlag in a rural population. *Chronobiol. Int.* **28**, 771–778 (2011).
41. Zordan, M. A. & Sandrelli, F. Circadian clock dysfunction and psychiatric disease: could fruit flies have a say? *Front. Neurol.* **6**, 80 (2015).
42. Logan, R. W. *et al.* Chronic stress induces brain region-specific alterations of molecular rhythms that correlate with depression-like behavior in mice. *Biol. Psychiatry* **78**, 249–258 (2015).
43. Zhan, S., Merlin, C., Boore, J. L. & Reppert, S. M. The monarch butterfly genome yields insights into long-distance migration. *Cell* **147**, 1171–1185 (2011).
44. Richards, S. *et al.* The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**, 949–955 (2008).
45. Weinstock, G. M. *et al.*; Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**, 931–949 (2006).


Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the members of the Tessmar-Raible, Raible and von Haeseler groups for discussions, and S. Bannister and F. Raible for comments on the manuscript. The research leading to these results has received funding from the research platform ‘Rhythms of Life’ of the University of Vienna to K.T.-R., T.H. and A.v.H., the FWF (<http://www.fwf.ac.at/>) START award (#AY0041321), the HFSP (<http://www.hfsp.org/>) research grant (#RGY0082/2010), the ERC (FP7/2007-2013)/ERC Grant Agreement 337011 to K.T.-R. and the DFG grant (#HE5398/4-1) to F.H. T.S.K. was supported by the Vienna International Postdoctoral Program for Molecular Life Sciences (VIPS), L.-T.N. and A.v.H. by the University of Vienna Initiativkolleg I059-N. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions T.S.K., K.T.-R. and A.v.H. conceived and designed the study and interpreted the data; T.S.K. carried out field sampling, chronobiological experiments, (super-) scaffolding, genetic/QTL mapping and population genomics; L.-T.N., T.N. and T.S.K. carried out the assembly to contigs; T.S.K. and T.N. carried out the assembly filtering; T.S.K. and F.J.S. carried out gap closing and repeated edge removal; T.S.K., F.J.S. and H.V. carried out RNA-seq; H.V. generated the cDNA library; T.S.K. and D.S. carried out genome annotation; T.S.K. and T.N. carried out analysis of genome completeness; T.S.K. and A.Z. carried out chromosome homology and synteny comparisons; A.J.B. carried out the estimation of linkage disequilibrium; D.S. and T.S.K. carried out the SNP effects and GO term analysis; B.P., T.S.K. and S.D. carried out *Cma-CaMKII.1* analyses; M.P. and F.H. carried out the minigene assay; T.H. contributed material; T.S.K. and K.T.-R. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.T.-R. (kristin.tessmar@mfpl.ac.at) or T.S.K. (kaiser@evolbio.mpg.de).

Reviewer Information Nature thanks M. Rosbash and the other anonymous reviewer(s) for their contribution to the peer review of this work.

 This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Animal culture and light regimes. The *C. marinus* laboratory stocks were bred according to Neumann¹, care was provided by the MFPL aquatic facility. Briefly, *C. marinus* were kept in 20 × 20 × 5 cm plastic containers with sand and natural seawater diluted to 15‰ with desalted water, fed diatoms (*Phaeodactylum tricornutum*, strain UTEX 646) in early larval stages and nettle powder in later stages. Temperature in the climate chambers was set to 20 °C and the light–dark cycle was 12:12 (except where noted differently). Moonlight was simulated with an incandescent flashlight bulb (about 1 lx), which was switched on all night for four successive nights every 30 days.

Genome assembly. The genome assembly process (Extended Data Fig. 9a) was based on three sequencing libraries (Supplementary Table 10): a 0.2-kb insert library was prepared from a single adult male of the Jean laboratory strain (established from field samples taken at St. Jean-de-Luz, France, in 2007; >12 generations in the laboratory), which was starved and kept in seawater with penicillin (60 units per ml), streptomycin (60 µg ml⁻¹) and neomycin (120 µg ml⁻¹) during the last 2 weeks of development. DNA was extracted with a salting-out method⁴⁶, sheared on a Covaris S2 sonicator (frequency sweeping mode; 4 °C; duty cycle, 10%; intensity, 7; cycles per burst, 300; microTUBE AFA fibre 6 × 16 mm; 30 s) and prepared for Illumina sequencing with standard protocols. A 2.2-kb and a 7.6-kb insert library were prepared from a polymorphic DNA pool of >300 field-caught Jean adult males by Eurofins MWG Operon (Ebersberg, Germany) according to the manufacturer's protocol. Each library was sequenced in one lane of an Illumina HiSeq2000 with 100-bp paired-end reads at the Next Generation Sequencing unit of the Vienna Biocenter Core Facilities (<http://vbcf.ac.at>).

Reads were filtered for read quality, adaptor and spacer sequences with cutadapt⁴⁷ (–b –n 3 –e 0.1 –O 8 –q 20 –m 13) and duplicates were removed with fastq-mcf from ea-utils⁴⁸ (–D 70). Read pairs were interleaved with ngm-utils⁴⁹, leaving only paired reads. Contamination with human DNA found in the 0.2-kb library was removed by deleting reads matching the human genome at a phred-scaled quality score ≥ 20 (alignment with BWA⁵⁰).

Assembly into contigs with Velvet⁵¹ (scaffolding disabled; 57-bp kmers as determined by VelvetOptimiser⁵²) was based solely on the less polymorphic 0.2-kb library. About 600 remaining adaptor sequences at the ends of assembled contigs were trimmed with cutadapt (–O 8 –e 0.1 –n 3). For assembly statistics see Supplementary Table 11.

Scaffolding of the contigs was based on all three libraries and performed with SSPACE⁵³ in two iterations, that is, scaffolds from the first round were scaffolded again. Using different parameters in the iterations (Supplementary Table 12) allowed different connections to be made and thus increased scaffold connectivity (Supplementary Table 13). The effect is probably owing to the polymorphic nature of the 2.2-kb and 7.6-kb libraries; it results in a 'population-consensus most common arrangement of the scaffolds'. The iterative scaffolding process was performed with and without applying a size cut-off excluding contigs <1 kb, resulting in two independent assemblies (CLUMA_0.3 and CLUMA_0.4; see Extended Data Fig. 9a), which differed in overall connectivity and sequence content (Supplementary Table 11), but also in the identity and structure of the large scaffolds. In order to combine both connectivity and sequence content, and in order to resolve the contradictions in the structure of the largest scaffolds, the two assemblies were compared and reconciled in a manual super-scaffolding process, as detailed in Supplementary Method 1. Briefly, the overlap of scaffolds from the two assemblies was tested with BLAST searches and represented in a graphical network structure. Scaffolds with congruent sequence content in both assemblies would result in a linear network, whereas scaffolds with contradictory sequence content would result in branching networks. At the same time, both assemblies were subject to genetic linkage mapping based on genotypes obtained from restriction-site-associated DNA sequencing (RAD sequencing) of a published mapping family⁶ (Supplementary Method 2). The resulting genetic linkage information served to resolve the branching networks into the longest possible unambiguous linear sub-networks with consistent genetic linkage information (see scheme A in Supplementary Method 1). Finally, the structure of the resulting super-scaffolds was coded in YAML format and translated into DNA sequence with Scaffoldr⁵⁴, resulting in 75 mapped super-scaffolds.

The remaining small and unmapped scaffolds were filtered for fragments of the mitochondrial genome, the histone gene cluster and 18S/28S ribosomal rDNA gene cluster, which were assembled separately (Supplementary Method 3; Extended Data Fig. 10). Unmapped scaffolds were also filtered for obvious contamination from other species (Supplementary Method 3). The degree to which the remaining unmapped scaffolds are leftover polymorphic variants of parts of the mapped super-scaffolds was estimated by blasting the former against the latter (Supplementary Method 3 and Supplementary Table 14).

All scaffolds were subject to gap closing with GapFiller⁵⁵ and repeated edges, that is, gaps with almost identical sequences at both sides that are generally not closed because of genetic polymorphisms, were assessed and if possible removed with a custom script (Supplementary Method 4; code available supplied as Source Data File).

The final assembly CLUMA_1.0 was submitted under project PRJEB8339 (75 mapped scaffolds; 23,687 unmapped scaffolds ≥ 100 bp). The assembly and further information can also be obtained from *ClunioBase* (<http://cluniobase.cibiv.univie.ac.at>).

Reconstruction of chromosomes and QTL analysis. Genetic linkage information for the final 75 super-scaffolds was obtained by repeating read mapping to genotype calling for the RAD sequencing experiment as described above (Supplementary Method 2), but now with assembly CLUMA_1.0 as a reference. This allowed us to place and orient super-scaffolds along the genetic linkage map (Fig. 1a and Extended Data Fig. 2). The positions of the recombination events within a scaffold were approximated as the middle between the positions of the two RAD markers between which the marker pattern changed from one map location to the next. The published genetic linkage map was refined and revised (Supplementary Method 5 and Extended Data Fig. 2). Based on the refined linkage map, QTL analysis of the published mapping family was repeated as described⁶ (Supplementary Table 4 and Supplementary Note 5). Using the correspondence between the reference assembly and the genetic linkage map, we were able to directly identify the genomic regions corresponding to the confidence intervals of the QTLs (Fig. 1 and Extended Data Fig. 5a, b).

Transcript sequencing. Assembled transcripts of a normalized cDNA library of all life stages and various *C. marinus* strains (454 sequencing) were available from previous experiments and RNA sequencing data was available for Jean strain adults (Illumina sequencing). Furthermore, specifically for genome annotation, RNA from 80 third instar larvae from the Jean and Por laboratory strains each was prepared for RNA sequencing according to standard protocols (Supplementary Method 6). Each sample was sequenced on a single lane of an Illumina HiSeq 2000. All transcript reads were submitted to the European Nucleotide Archive (ENA) under project PRJEB8339.

For the adult and larval RNA sequencing data, raw reads were quality checked with fastqc⁵⁶, trimmed for adaptors quality with cutadapt⁴⁷ and filtered to contain only read pairs using the interleave command in ngm-utils⁴⁹. Reads were assembled separately for larvae and adults with Trinity⁵⁷ (path_reinforcement_distance: 25; maximum paired-end insert size: 1,500 bp; otherwise default parameters).

Genome annotation. Automated annotation was performed with MAKER2⁵⁸. Repeats were masked based on all available databases in repeatmasker. MAKER2 combined evidence from assembled transcripts (see above), mapped protein data sets from *Culex quinquefasciatus* (CpipJ1), *Anopheles gambiae* (AgamP3), *Drosophila melanogaster* (BDGP5), *Danaus plexippus* (DanPle_1.0), *Apis mellifera* (Amel4.0), *Tribolium castaneum* (Tcas3), *Strigamia maritima* (Smar1) and *Daphnia pulex* (Dappu1) and *ab initio* gene predictions with AUGUSTUS⁵⁹ and SNAP⁶⁰ into gene models. AUGUSTUS was trained for *C. marinus* based on assembled transcripts from the normalized cDNA library. SNAP was run with parameters for *A. mellifera*, which had the highest congruence with known *C. marinus* genes in preliminary trials (Supplementary Method 7). MAKER was set to infer gene models from all evidence combined (not transcripts only) and gene predictions without transcript evidence were allowed. Splice variant detection was enabled, single-exon genes had to be larger than 250 bp and intron size was limited to a maximum of 10 kb.

All gene models within the QTL confidence intervals, as well as all putative circadian clock genes and light receptor genes were manually curated: exon–intron boundaries were corrected according to transcript evidence (approximately 500 gene models), chimeric gene models were separated into the underlying individual genes (approximately 100 gene models separated into around 300 gene models) and erroneously split gene models were joined (approximately 15 gene models). Finally, this resulted in 21,672 gene models, which were given IDs from CLUMA_CG000001 to CLUMA_CG021672 ('CLUMA' for *Clunio marinus*, following the controlled vocabulary of species from the UniProt Knowledgebase; CG for 'computed gene'). Splice variants of the same gene (detected in 752 gene models) were identified by the suffix '-RA', '-RB' and so on, and the corresponding proteins by the suffix '-PA', '-PB' and so forth.

Gene models were considered as supported if they overlapped with mapped transcripts or protein data (Supplementary Table 1). Gene counts for *D. melanogaster* were retrieved from BDGP5, version 75.546 and for *A. gambiae* from AgamP3, version 75.3. The putative identities of the *C. marinus* gene models were determined in reciprocal BLAST searches, first against UniProtKB/Swiss-Prot (8,379 gene models assigned) and if no hit was found, second against the non-redundant protein sequences (nr database) at NCBI (1,802 additional genes assigned). Reciprocal best hits with an *e* value < 1 × 10⁻¹⁰ were considered putative orthologues (termed 'putative gene X'), non-reciprocal hits with the same *e* value were considered paralogues (termed 'similar to'). All remaining gene models

were searched against the PFAM database of protein domains (111 gene models assigned; termed 'gene containing domain X'). If still no hit was found, the gene models were left unassigned ('NA').

Synteny comparisons. Genome-wide synteny between the *C. marinus*, *D. melanogaster* and *A. gambiae* genomes was assessed based on reciprocal best BLAST hits (e value $< 10 \times 10^{-10}$) between the three protein data sets (Ensembl Genomes, Release 22, for *D. melanogaster* and *A. gambiae*). Positions of pairwise orthologous genes were retrieved from the reference genomes (BDGP5, AgamP3 and CLUMA_1.0) and plotted with Circos⁶¹. *C. marinus* chromosome arms were delimited based on centromeric and telomeric signatures in genetic diversity and linkage disequilibrium (Extended Data Fig. 3c and Supplementary Table 3; for data source see 'strain re-sequencing' below). Homologues for *C. marinus* chromosome arms were assigned based on enrichment with putative orthologous genes from specific chromosome arms in *D. melanogaster* and *A. gambiae* (Extended Data Figs 3, 4 and Supplementary Table 3). Additionally, for the 5,388 detected putative 1:1 orthologues (*C. marinus*:*D. melanogaster*:*A. gambiae*), microsynteny was assessed by testing if all pairs of directly adjacent genes in one species were also directly adjacent in the other species. The degree of microsynteny was then calculated as the fraction of conserved adjacencies among all pairs of adjacent genes. From this fraction the relative levels of chromosomal rearrangements in the evolutionary lineage leading to *C. marinus* were estimated (Supplementary Note 3 and Extended Data Fig. 4).

Strain re-sequencing. Genetic variation in five *C. marinus* strains (Extended Data Fig. 1) was assessed based on pooled-sequencing data from field-caught males from the strains of St. Jean-de-Luz (Jean; Basque Coast, France; sampled in 2007; $n = 300$), Port-en-Bessin (Por; Normandie, France; 2007; $n = 300$), as well as Vigo (Spain; 2005; $n = 100$), Helgoland (He; Germany; 2005; $n = 300$) and Bergen (Ber; Norway; 2005; $n = 100$). Samples from Vigo and Bergen, were provided by D. Neumann and C. Augustin, respectively. For each strain we chose the largest available number of individuals to obtain the best possible resolution of allele frequencies. Females are not available, because they are virtually invisible in the field. For an overview of the experimental procedure, see Extended Data Fig. 9b. DNA was extracted with a salting-out method⁴⁶ from sub-pools of 50 males, the DNA pools were mixed at equal DNA amounts, sheared and prepared as described above and sequenced on four lanes of an Illumina HiSeq2000 with paired-end 100-bp reads (Ber and Vigo combined in one lane, distinguished by index reads). All reads were submitted to the European Nucleotide Archive (ENA) under project PRJEB8339. Sequencing reads were filtered for read quality and adaptor sequences with cutadapt⁴⁷ ($-b -n 2 -e 0.1 -O 8 -q 13 -m 15$), interleaved with ngm-utils⁴⁹ and duplicates were removed with fastq-mcf from ea-utils⁴⁸ ($-D 70$). Reads were aligned to the mapped super-scaffolds of assembly CLUMA_1.0 with BWA⁵⁰ (aln and sampe; maximal insert size (bp): $-a 1500$).

Detection of re-arrangements. Based on the unfiltered alignments, the samples from Por and Jean were screened for genomic inversions and indels relative to the reference sequence with the multi-sample version of DELLY⁶². Paired-end information was only considered if the mapping quality was high ($q \geq 20$) (see also Supplementary Note 3).

Population genomic analysis of the timing strains. For population genomic analysis (Extended Data Fig. 9b), the alignments of the pool-sequencing (pool-seq) data from Vigo, Jean, Por, He and Ber were filtered for mapping quality ($q \geq 20$), sorted, merged and indexed with SAMtools⁶³. Reads were re-aligned around indels with the RealignerTargetCreator and the IndelRealigner in GATK⁶⁴. The resulting coverage per strain is given in Supplementary Table 5.

For identification of SNPs, a pileup file was created with the mpileup command of SAMtools⁶³. Base Alignment Quality computation was disabled ($-B$); instead, after creating a synchronized file with the mpileup2sync script in PoPoolation2⁶⁵, indels that occurred more than ten times were masked (including 3 bp upstream and downstream) with the identify-indel-regions and filter-sync-by-gtf scripts of PoPoolation2. F_{ST} values were determined with the fst-sliding script of PoPoolation2, applying a minimum allele count of 10 (so that any false-positive SNPs resulting from the remaining unmasked indels were effectively excluded) and a minimum coverage of $40 \times$ for the comparison between Por and Jean or $10 \times$ for the comparison of all five strains. F_{ST} was calculated at a single base resolution, as well as in windows of 5 kb (step size, 1 kb). Individual SNPs were only considered for further analyses or plotted if they were significantly differentiated as assessed by Fisher's exact test (fisher-test in PoPoolation2).

Average genome-wide genetic differentiation between timing strains, as obtained by averaging over 5-kb sliding-windows, was compared to the respective timing differences and geographic distances (see Supplementary Table 8) in Mantel tests (Pearson's product moment correlation; 9,999 permutations), as implemented in the vegan package in the R statistical programming environment (ref. 66). Geographic distances and circadian timing differences were determined as described previously⁶⁷ (see Supplementary Table 8). For determination

of lunar timing differences when comparing lunar with semilunar rhythms see Supplementary Note 6. In order to find genomic regions for which genetic differentiation is correlated with the timing differences between strains, the Mantel test was then applied to 5-kb genomic windows every 1 kb along the reference sequence. 5 kb is roughly the average size of a gene locus in *C. marinus*. Windows with a correlation coefficient of $r \geq 0.5$ were tested for significance (999 permutations). For each genomic position the number of overlapping significantly correlated 5-kb windows was enumerated, resulting in a correlation score (CS; ranging from 0 to 5).

Genetic diversity, measured as Watterson's theta (θ_W), for each strain was assessed with PoPoolation1.1.2 (ref. 68) in 20-kb windows with 10-kb steps. In order to save computing time, the pileup files of Jean, Por and He were linearly downscaled to $100 \times$ coverage with the subsample-pileup script ('fraction' option), positions below $100 \times$ coverage were discarded. Indel regions were excluded (default in PoPoolation 1.1.2) and a minimum of 66% of a sliding window needed to be covered. SNPs were only considered in θ_W calculations if present ≥ 2 times, leading to slight inconsistencies in θ_W estimates between strains due to differing coverage, but not affecting diversity comparisons within strains.

Linkage disequilibrium between the SNPs was determined for the Por and Jean strains with LDx⁶⁹, assuming physical linkage between alleles on the same read or read pairs. r^2 was determined by a maximum likelihood estimator, minimum and maximum read depths corresponded to the 2.5% and 97.5% coverage depths for each population (Jean, 111–315; Por, 98–319), total insert distance was limited to 600 bp, minimum phred-scaled base quality was 20, minimum allele frequency was 0.1 and a minimum coverage per pair of SNPs was 11. SNPs were binned by their physical distance for the plots (0–200 bp, 200–400 bp, 400–600 bp), with the mean value plotted.

Finally, small indels (< 30 bp) in the Por and Jean strains were detected with the UnifiedGenotyper ($-glm INDEL$) in GATK⁶⁴ for positions with more than $20 \times$ coverage. Genetic differentiation for indels was calculated with the classical formula $F_{ST} = (H_T - H_S)/H_T$, where H_S is the average expected heterozygosity according to Hardy-Weinberg Equilibrium (HWE) in the two subpopulations and H_T is the expected heterozygosity in HWE of the hypothetical combined total population. If more than two alleles were present, only the two most abundant alleles were considered in the calculation of F_{ST} .

Assessment of candidate genes. Gene models from the automated annotation were considered candidate genes, if they fulfilled the following criteria. (1) The gene was located within the reference sequence corresponding to the QTL confidence intervals as determined for the Por and Jean strains. (2) The gene contained a strongly differentiated SNP or small indel or it was directly adjacent to such a SNP or small indel ($F_{ST} \geq 0.8$ for Por versus Jean, that is, the strains used in QTL mapping). This resulted in a preliminary list of 133 genes based on the comparison between Por and Jean (Supplementary Table 6). These candidate genes were narrowed down based on their overlap with genomic 5-kb windows, for which genetic differentiation between five European timing strains correlated with their timing differences (Fig. 1a, Extended Data Fig. 5a, b and Supplementary Table 9).

The location and putative effects of the SNPs and indels relative to the gene models were assessed with SNPeff⁷⁰ ($-ud 0$, otherwise default parameters; Extended Data Fig. 5c, d and Supplementary Tables 6, 9).

For Gene Ontology (GO) term analysis, all *C. marinus* gene models with putative orthologues in the UniProtKB/Swiss-Prot and non-redundant protein sequences (nr) databases based on reciprocal best BLAST hits (see above) were annotated with the GO terms of their detected orthologues (6,837 gene models). Paralogues were not annotated. The enrichment of candidate SNPs and indels ($F_{ST} \geq 0.8$ between Por and Jean) in specific GO terms was tested with SNP2GO⁷¹ (min.regions = 1, otherwise default parameters). Hyper-geometric sampling was applied to test if individual genes of a GO term or a whole pathway of genes are enriched for SNPs (Supplementary Table 7).

Molecular characterization of CaMKII.1. RNA-seq data of the Por and Jean strains for *CaMKII.1* were obtained from the larval RNA sequencing experiment described above. Besides four assembled full-length transcripts (RA-RD) from RNA-seq and assembled EST libraries, additional partial transcripts (RE-RO) were identified by PCR amplification (for PCR primers see Supplementary Table 15), gel extraction (QIAquick Gel Extraction Kit, Qiagen), cloning with the CloneJET PCR Cloning Kit (Thermo Scientific) and Sanger sequencing with pJET1.2 primers (LGC Genomics & Microsynth). cDNA was prepared from RNA extracted from third instar larvae of the Por and Jean laboratory strains (RNA extraction with RNeasy Plus Mini Kit, Qiagen; reverse transcription with QuantiTect Reverse Transcription Kit, Qiagen).

qPCR was performed with variant-specific primers and actin was used as a control gene (Supplementary Table 16). cDNA was obtained from independent pools of 20 third instar larvae of the Por and Jean strains. Sample size was ten pools per strain to cover different time points during the day and to test for reproducibility

(two samples each at zeitgeber times 0, 4, 8, 16 and 20; for one Por sample extraction failed; RNA extraction and reverse transcription as above). qPCR was performed with Power SYBR Green PCR Master Mix on a StepOnePlus Real Time System (both Applied Biosystems). Fold-changes were calculated according to ref. 72 in a custom excel sheet. The assumption of equal variance was violated for the RD comparison (*F*-test) and the assumption of normal distribution was violated for the data of RA and RC in the Por strain (Shapiro–Wilk normality test), possibly reflecting circadian effects in the samples from different times of day. Thus, expression differences were assessed for significance in a two-tailed Wilcoxon rank-sum test (wilcox.test in R⁶⁶). Holm correction⁷³ was used for multiple testing (default in p.adjust function of R).

CaMKII.1 minigenes. PCR fragments containing the CaMKII.1 linker region (exons 10–15) were amplified from genomic Por or Jean DNA, respectively, with primers CaMKII-Sc61-F-344112 and CaMKII-Sc61-R-351298 (Supplementary Table 15), cloned with the CloneJET PCR Cloning Kit (Thermo Scientific), transferred into the pcDNA3.1+ vector using NotI and XbaI (Thermo Scientific). These constructs were transfected into *D. melanogaster* S2R+ cells and RNA was prepared 48 h after transfection. After DNase digestion, isoform expression was analysed by radioactive, splicing-sensitive RT-PCR (primers in Supplementary Table 17) and phosphorimager quantification as described⁷⁴. Identity of isoforms is based on size and sequencing of PCR products. To test for reproducibility, there were seven biological replicates (raw data in Supplementary Table 18). As the assumptions of equal variance (*F*-test) and normal distribution of data (Shapiro–Wilk normality test) were not violated, the significance of expression differences was assessed in unpaired, two-sided two-sample *t*-tests. Holm correction⁷³ was used for multiple testing (default in p.adjust function of R). S2R+ cells were obtained from the laboratory of S. Sigrist, regularly authenticated by morphology and routinely tested for absence of mycoplasma contamination. The entire experiment was reproduced several months later with three biological replicates (raw data in Supplementary Table 18).

S2 cell luciferase assay. Firefly luciferase is driven from a period 3X69 promoter under control of the CLOCK and CYCLE protein^{19,21}. The *D. melanogaster* pAc-clk construct was obtained from F. Rouyer, pCopia–Renilla luciferase and period 3X69–luc reporter constructs from M. Rosbash, a [Ca²⁺]-independent mouse CaMKII^{T286D} was provided by M. Mayford. The CaMKII inhibitor KN-93 was purchased from Abcam (#ab120980).

C. marinus Cyc, *C. marinus* Clk and *C. marinus* CaMKII.1–RD were cloned into the pAc5.1/V5–His A plasmid (Invitrogen) with stop codons before the tag. The Q5 Site-Directed Mutagenesis Kit (NEB) was used to make kinase-dead and [Ca²⁺]-independent versions of *C. marinus* CaMKII.1–RD (for primers, see Supplementary Table 17).

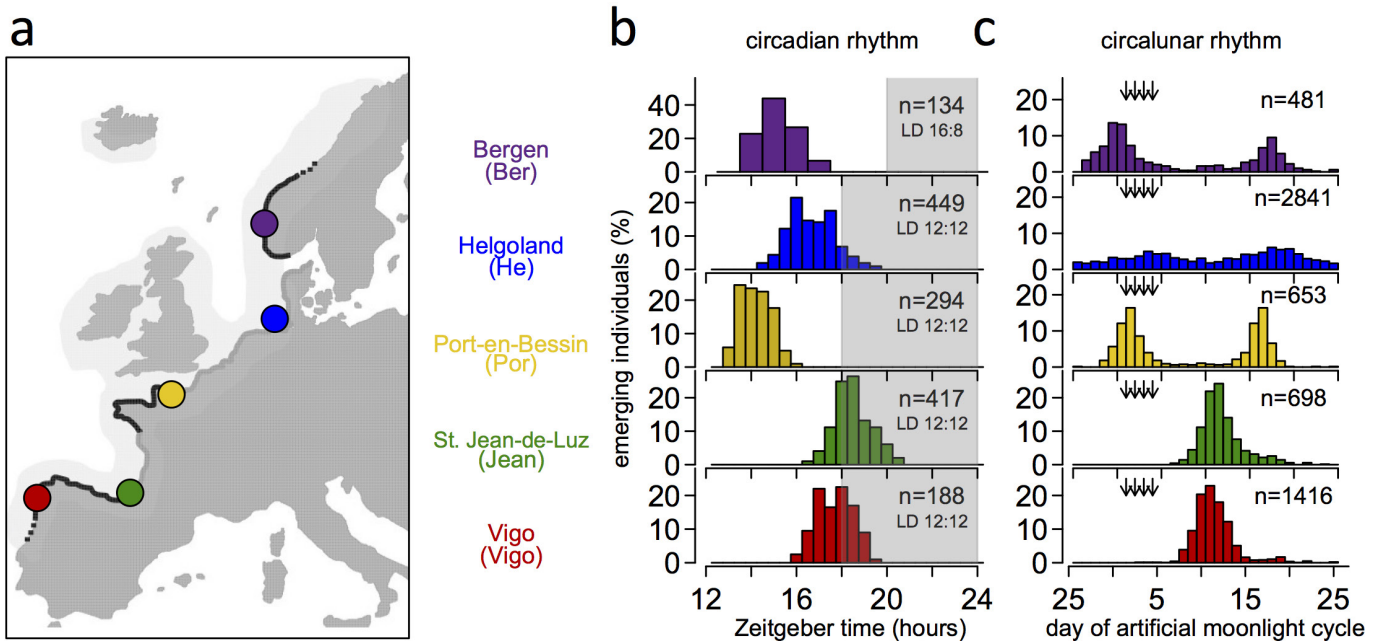
D. melanogaster S2 cells (Invitrogen) were cultured at 25 °C in Schneider's *D. melanogaster* medium (Lonza) supplemented with fetal bovine serum (FBS, 10%, heat-inactivated), penicillin (100 U ml^{−1}), streptomycin (100 µg ml^{−1}) and 2 mM L-glutamine; Sigma). Cells were seeded into 24-well plates (800,000 cells per well) and transfected with Effectene transfection reagent (Qiagen) according to the manufacturer's instructions. Experiment with mouse [Ca²⁺]-independent CaMKII: 25 ng pCopia–Renilla, 10 ng period 3X69–luc, 0.5 ng *D. melanogaster* pAc-clk, 200 ng mouse pAc–CaMKII^{T286D}. Experiment with CaMKII inhibitor KN-93: 25 ng pCopia–Renilla, 10 ng period 3X69–luc, 0.5 ng *D. melanogaster* pAc-clk, various amounts of KN-93. Experiment with *C. marinus* genes: 25 ng pCopia–Renilla, 10 ng period 3X69–luc, 100 ng *C. marinus* pAc–cyc, 100 ng *C. marinus* pAc-clk, 200 ng *C. marinus* CaMKII.1–RD^{K42R} or 200 ng *C. marinus* CaMKII.1–RD^{T286D}. In all experiments, the transfection mix was filled up with empty pAc5.1/V5–His A vector to a total of 435 ng DNA per well. After 48 h, cells were washed with PBS and lysed with Passive Lysis Buffer (Promega). Luciferase activities were determined on a Synergy H1 plate reader (Biotek) using a Dual-Luciferase Reporter Assay System (Promega). For each biological replicate three independent cell lysates were measured and their mean value determined. Firefly luciferase activity was normalized to Renilla luciferase activity and values were normalized to controls transfected with *D. melanogaster* pAc-clk or *C. marinus* pAc-clk and *C. marinus* pAc–cyc, respectively. S2 cells (Invitrogen/Life Technologies, Cat.no. R690-07) were regularly authenticated by morphology and routinely tested for absence of mycoplasma contamination (Lonza MycoAlert). Sample size was chosen to test for reproducibility.

Circadian free-run experiments. For circadian free-run experiments, culture boxes of the Por, He and Jean strains were transferred from light–dark cycle (16:8) to constant dim light (light–light cycle, about 100 lx). Emerging adults were collected in 1-h intervals by a custom made *C. marinus* fraction collector (similar to those described in ref. 75) and counted once a day. Because collection was automated, the experimenter had no influence on the results and blinding was not necessary. As the circalunar clock restricts adult emergence to a few days, the circadian emergence rhythm can only be assessed over a few days. Several culture

boxes were transferred to a light–light cycle at different time points. The resulting emergence data were combined for each strain using the switch to a light–light cycle as a common reference point. We used the maximum number of available individuals. Free-running period was calculated as the mean interval between subsequent emergence peaks, weighting each peak by the number of individuals.

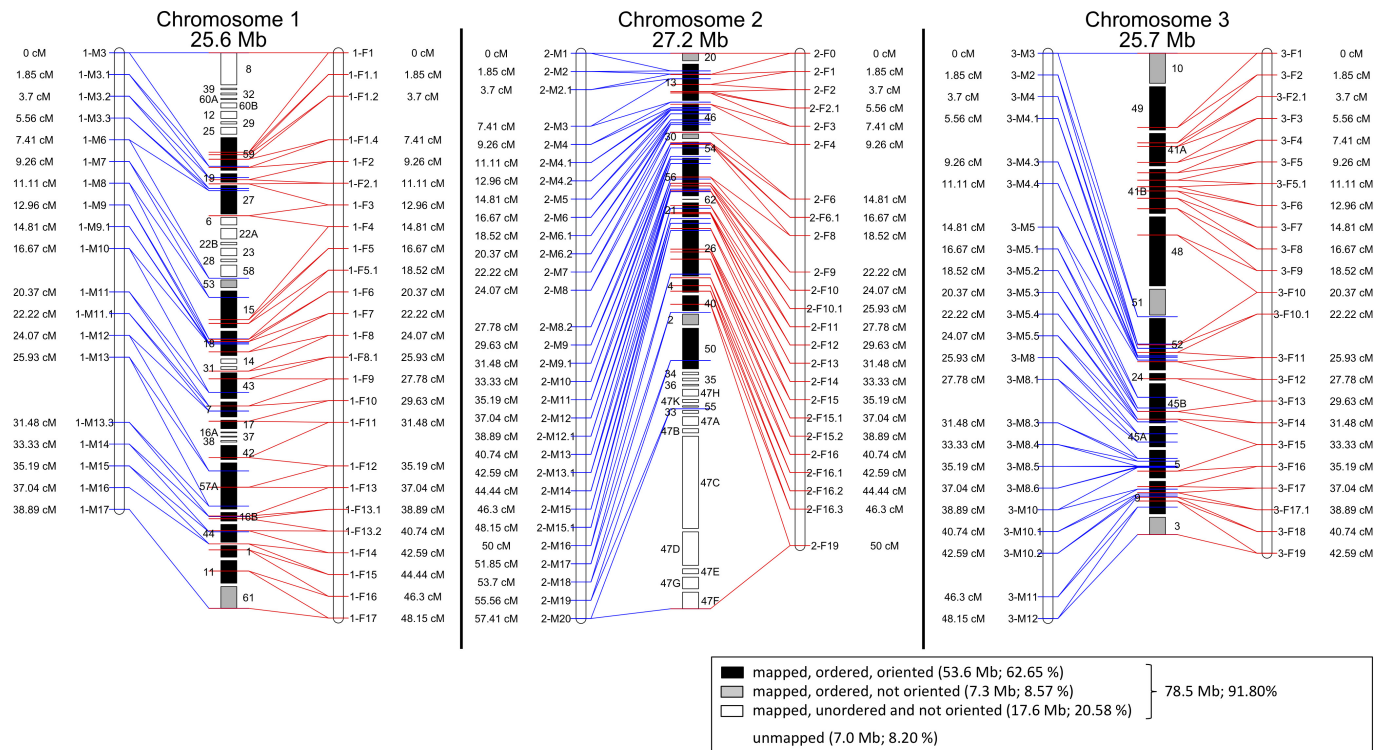
Data availability. All sequence data are deposited in the European Nucleotide Archive (ENA) under PRJEB8339. The reference genome is also on *ClunioBase* (<http://clunioibase.cibiv.univie.ac.at>). Machine readable super-scaffolding data and the computer source code for the removal of repeated edges are supplied as source data files.

46. Reineke, A., Karlovsky, P. & Zebitz, C. P. W. Preparation and purification of DNA from insects for AFLP analysis. *Insect Mol. Biol.* **7**, 95–99 (1998).
47. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
48. Aronesty, E. Command-line tools for processing biological sequencing data <http://code.google.com/p/ea-utils> (2011).
49. Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790–2791 (2013).
50. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
51. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
52. Gladman, S. & Seemann, T. VelvetOptimiser <http://bioinformatics.net.au/software/velvetoptimiser.shtml> (2012).
53. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
54. Barton, M. D. & Barton, H. A. Scaffolder—software for manual genome scaffolding. *Source Code Biol. Med.* **7**, 4 (2012).
55. Boetzer, M. & Pirovano, W. Toward almost closed genomes with GapFiller. *Genome Biol.* **13**, R56 (2012).
56. Andrews, S. FastQC Version 0.10.1. A Quality Control tool for High Throughput Sequence Data <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2012).
57. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
58. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
59. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** (Suppl 2), ii215–ii225 (2003).
60. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
61. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
62. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
63. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
64. McKenna, A. *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
65. Kofler, R., Pandey, R. V. & Schlötterer, C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435–3436 (2011).
66. Crawley, M. J. *The R Book*. (John Wiley & Sons Ltd, 2007).
67. Kaiser, T. S., Neumann, D., Heckel, D. G. & Berendonk, T. U. Strong genetic differentiation and postglacial origin of populations in the marine midge *Clunio marinus* (Chironomidae, Diptera). *Mol. Ecol.* **19**, 2845–2857 (2010).
68. Kofler, R. *et al.* PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* **6**, e15925 (2011).
69. Feder, A. F., Petrov, D. A. & Bergland, A. O. LDx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. *PLoS One* **7**, e48588 (2012).
70. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
71. Szikiba, D., Kapun, M., von Haeseler, A. & Gallach, M. SNP2GO: functional analysis of genome-wide association studies. *Genetics* **197**, 285–289 (2014).
72. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{−(ΔΔC_T)} method. *Methods* **25**, 402–408 (2001).
73. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
74. Preußner, M. *et al.* Rhythmic U2af26 alternative splicing controls PERIOD1 stability and the circadian clock in mice. *Mol. Cell* **54**, 651–662 (2014).
75. Honegger, H. W. An automatic device for the investigation of the rhythmic emergence pattern of *Clunio marinus*. *Int. J. Chronobiol.* **4**, 217–221 (1977).
76. Heimbach, F. *Semilunare und diurnale Schlüpfrythmen südeuropäischer und norwegischer Clunio-Populationen* (Diptera, Chironomidae) PhD thesis, Universität Köln (1976).
77. Friedrich, M. & Tautz, D. Evolution and phylogeny of the Diptera: a molecular phylogenetic analysis using 28S rDNA sequences. *Syst. Biol.* **46**, 674–698 (1997).
78. Zdobnov, E. M., *et al.* Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**, 149–159 (2002).

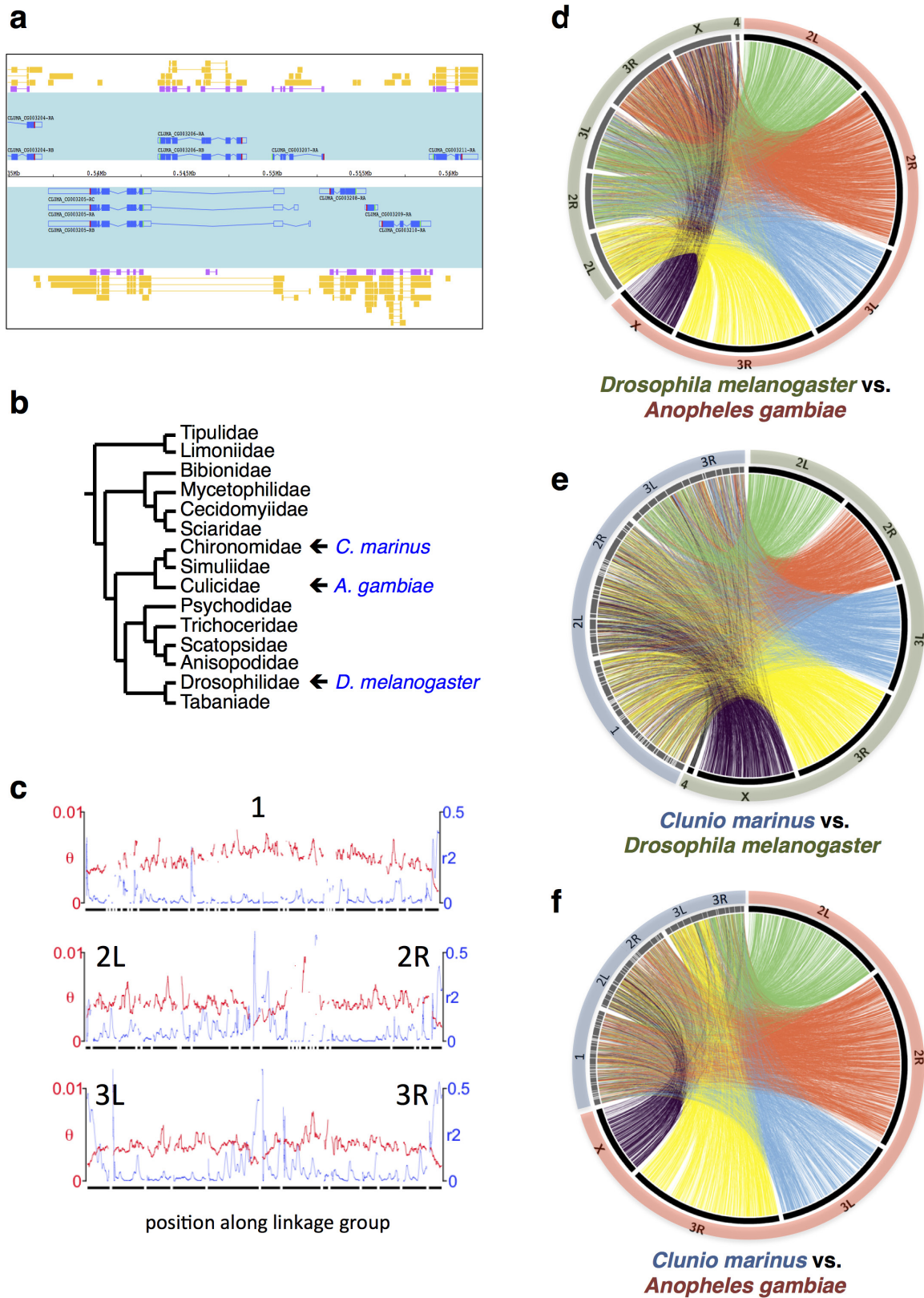


Extended Data Figure 1 | The biology of *Clunio marinus*. **a**, *C. marinus* is restricted to rocky shores (black lines), the localities differing in tidal regime (adapted from ref. 67). **b**, **c**, Local strains show corresponding genetic adaptations in their circadian⁶⁷ (**b**) and circalunar rhythms (**c**, He¹, Jean⁵). Timing was measured in the laboratory under artificial moonlight

(arrows in **c**) in a 30-day cycle and a light–dark cycle of 12:12 (He, Por, Jean, Vigo) or 16:8 (Ber). Seasonal differences in daily illumination duration do not affect circadian emergence peaks^{1,76}. Historically, for *C. marinus* ‘zeitgeber time 0’ is defined as the middle of the dark phase.



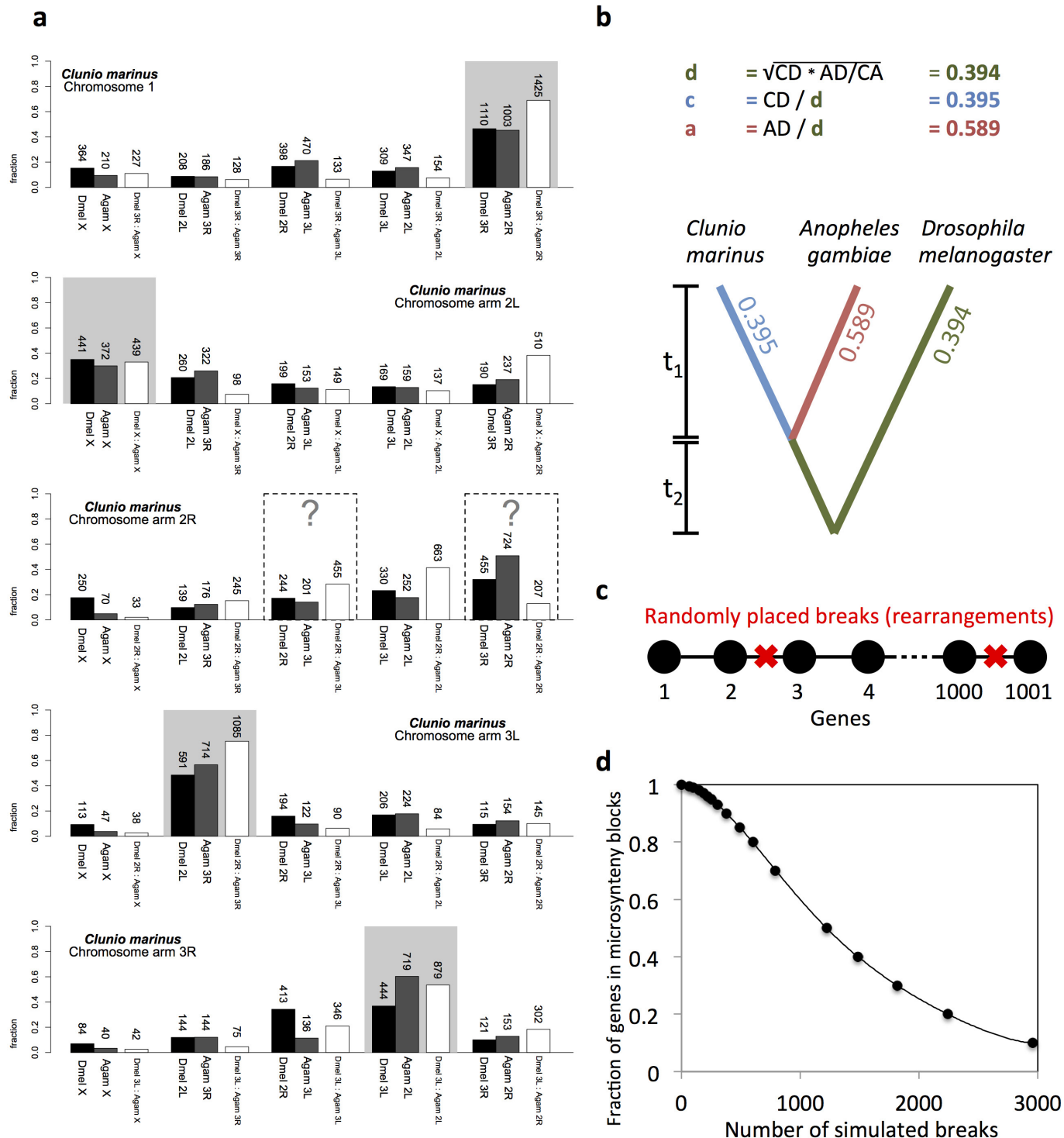
Extended Data Figure 2 | The reconstructed chromosomes of *C. marinus* based on the genetic linkage map. Left map, male informative markers. Right map, female informative markers. See Fig. 1a legend for further details.



Extended Data Figure 3 | *C. marinus* genome characterization.

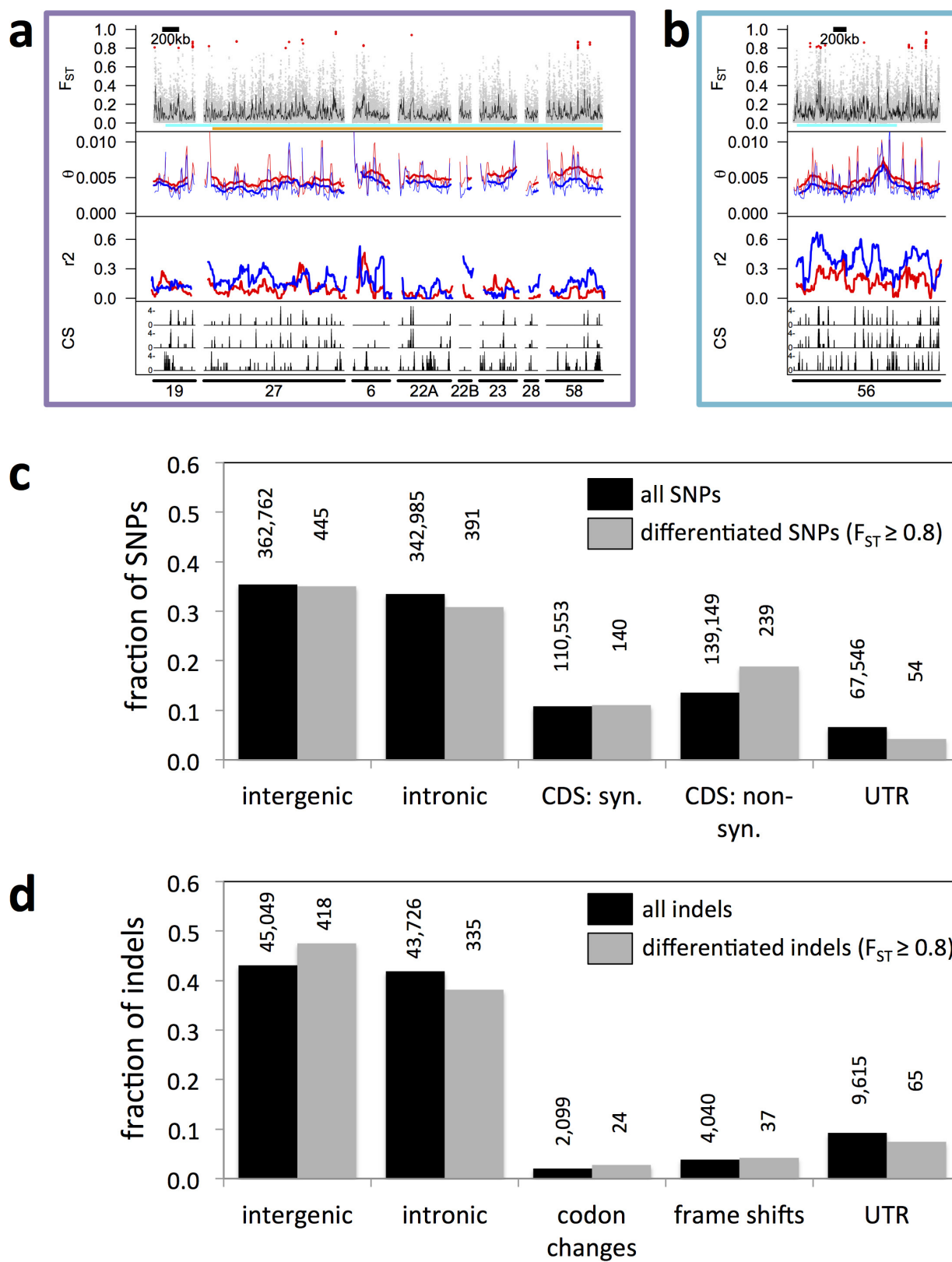
a, Representative genomic region with densely packed gene models (super-scaffold 1, 535–565 kb). Gene models are shown in blue on a turquoise background. Gene predictions (SNAP) are shown in purple. Transcript evidence is shown in yellow. **b**, Phylogenetic relationships of *C. marinus* to other Diptera (according to ref. 77). **c**, Genetic diversity

(θ ; red) and linkage disequilibrium (r^2 ; blue) of the Jean strain plotted for the three *C. marinus* linkage groups, revealing characteristic signatures of telomeres and centromeres. **d–f**, Synteny comparisons among the genomes of *C. marinus*, *A. gambiae* and *D. melanogaster* based on 5,388 1:1 orthologues.



Extended Data Figure 4 | Synteny analyses of *C. marinus* chromosome arms. **a**, Gene content of the *C. marinus* chromosome arms relative to the chromosome arms of *D. melanogaster* (black bars) and *A. gambiae* (grey bars). The very small chromosome 4 of *D. melanogaster* is neglected. Chromosome arms of *D. melanogaster* and *A. gambiae* are paired according to their published homology⁷⁸. For four of the chromosome arms of *C. marinus* the homologous arms in *D. melanogaster* and *A. gambiae* are identified (grey shading). For comparison, the conservation of the identified *D. melanogaster* and *A. gambiae* homologues to each other is given by plotting the gene content of the homologous *D. melanogaster* chromosome arm relative to the different chromosome arms of *A. gambiae* (white bars). The numbers of orthologous genes considered in

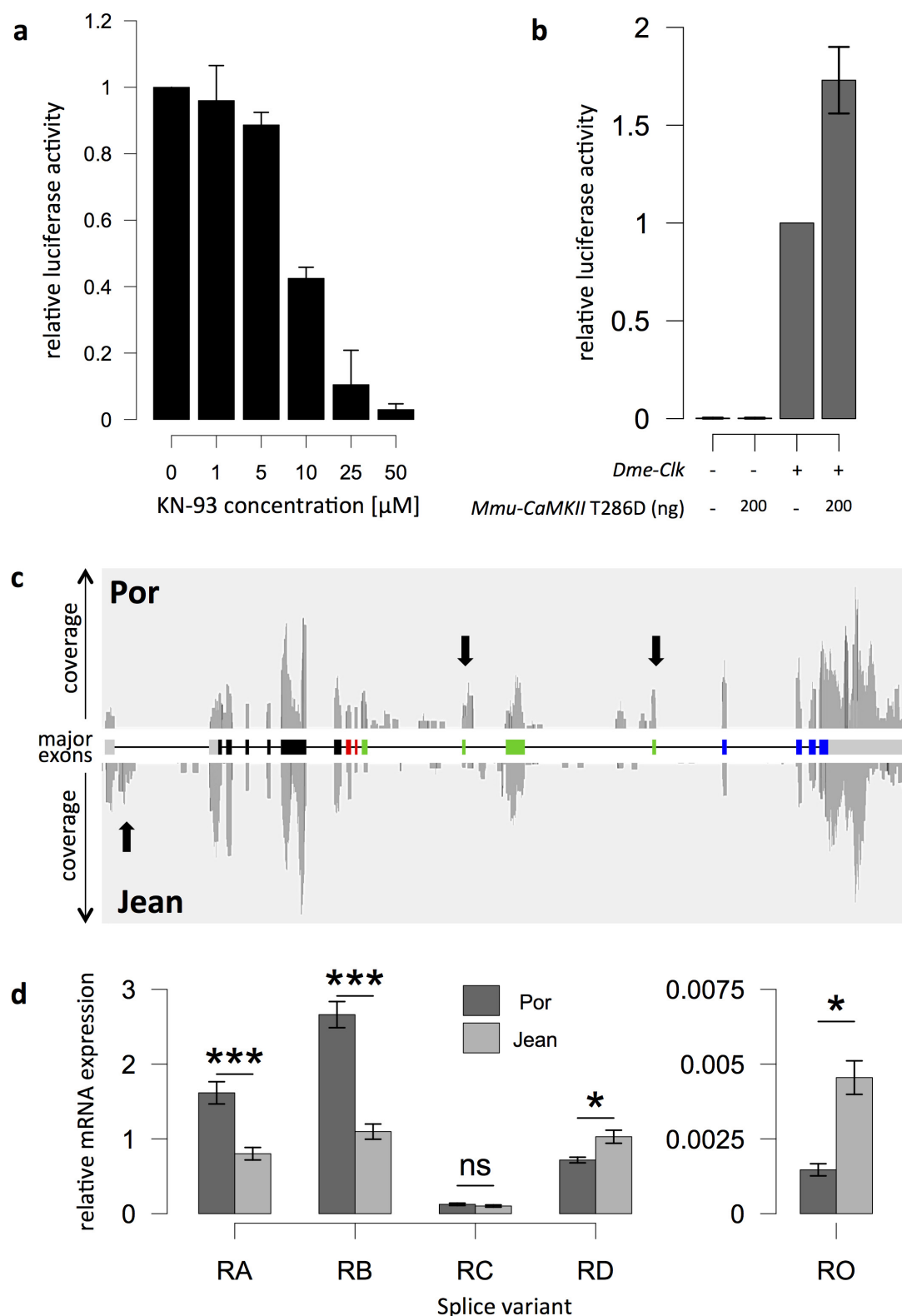
each comparison are given above the bars. For chromosome arm 2R of *C. marinus* the homologies are unclear. Possibly, chromosome arm 2R of *C. marinus* has undergone so many re-arrangements with other chromosome arms that it is no longer recognizable, which is consistent with complex polymorphic re-arrangements in this chromosome arm of *C. marinus* (see Supplementary Note 3). **b**, Microsynteny is analysed relative to *D. melanogaster* and *A. gambiae*, based on 5,388 1:1:1 orthologues. The fraction of genes in conserved microsynteny blocks is calculated and distributed along the phylogenetic tree. **c, d**, A simulation was used to estimate how many chromosomal re-arrangements are required to produce the observed degree of microsyntenic conservation (for details see Supplementary Note 3).



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Population genomic analysis of QTLs C1/L1 and C2 and genome-wide analysis of locations and putative effects of SNPs and indels. **a, b**, Population genomic analysis of QTLs C1/L1 and C2. Panels 1–3: Por versus Jean strains in blue and red, respectively, in panel 2 and 3. From top to bottom, panel 1, genetic differentiation (red dots, SNPs with $F_{ST} \geq 0.8$; grey dots, $F_{ST} < 0.8$; black line, average F_{ST} in 5-kb sliding windows). Panel 2, genetic diversity (θ) in 20-kb (thin line) and 200-kb (thick line) windows. Panel 3, linkage disequilibrium (r^2) for SNP pairs 0–600 bp apart in 100-kb windows (step size: 5 kb). Panel 4, correlation score (CS; 0–5) for genetic differentiation with circadian timing (top), circalunar timing (middle) and geographic distance (bottom) for five European *C. marinus* strains (Vigo, Jean, Por, He and Ber). Bottom numbers, scaffold IDs. See also Fig. 1. **c, d**, Locations and putative effects

of SNPs (**c**) and indels (**d**) with respect to the annotated gene models. The fractions of SNPs or indels in each category are compared for all SNPs and indels (black bars) versus differentiated SNPs and indels ($F_{ST} \geq 0.8$ between Por and Jean strains; grey bars). Absolute numbers are given above the bars. In gene models with several splice forms, SNPs and indels can have different effects, for example, ‘CDS, non-synonymous’ for one splice form and ‘intronic’ for another splice form. Therefore, the sum across locations is slightly larger than the actual numbers of SNPs and indels. ‘Codon changes’ are all codon insertions or deletions that do not result in frame shifts beyond the actual insertion/deletion site. CDS, coding sequence; syn., synonymous; non-syn., non-synonymous; UTR, untranslated region.

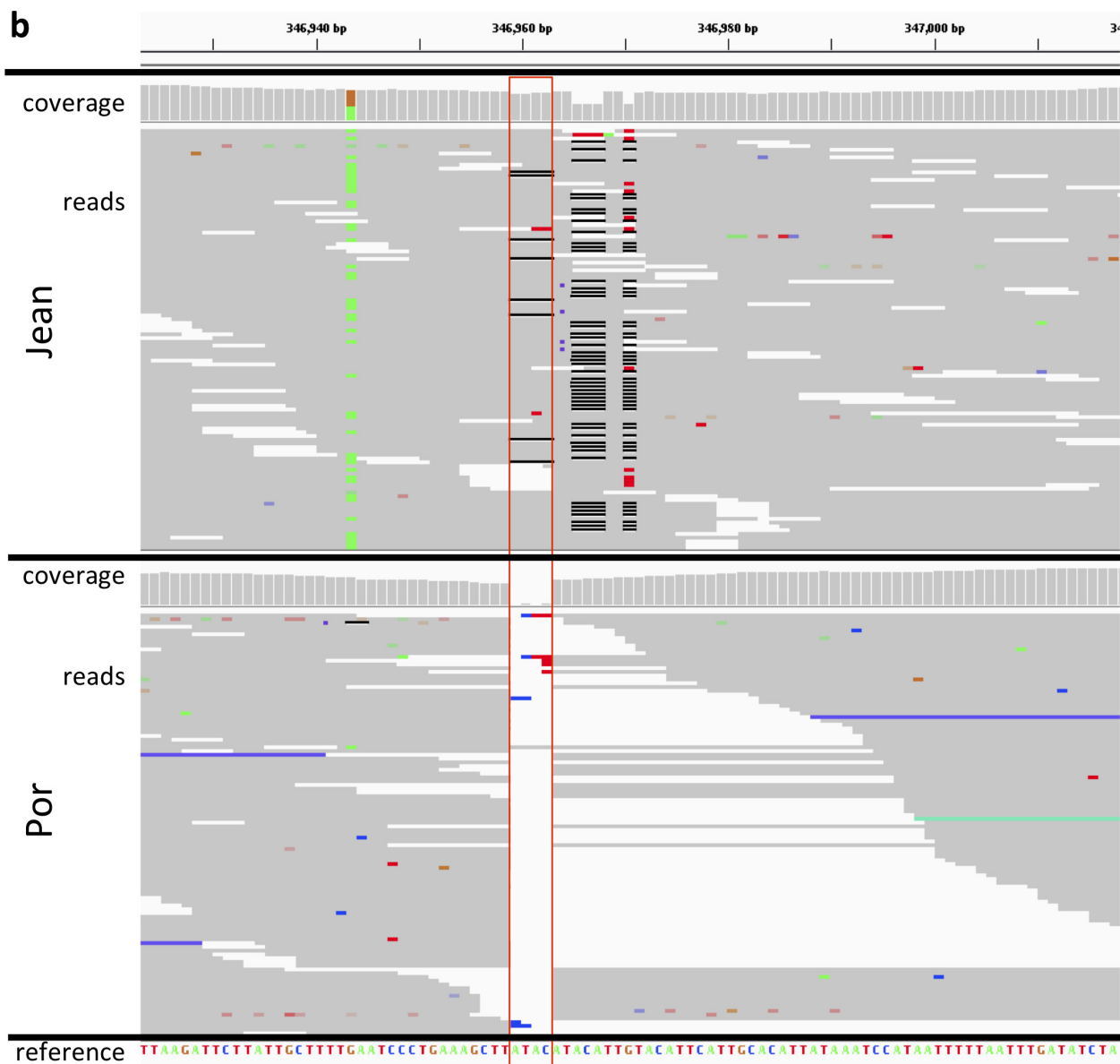


Extended Data Figure 6 | CaMKII regulates CLK/CYC transcriptional activity and exhibits strain-specific splice variants. **a**, Quantification of luciferase activity under the control of an artificial 3X69 E-box containing enhancer in *D. melanogaster* S2 cells. Increasing amounts of the CaMKII inhibitor KN-93 decrease luciferase activity in a concentration-dependent manner, providing evidence that endogenous CaMKII activity regulates the transcriptional activity of *D. melanogaster* CLOCK-CYCLE. **b**, Without co-transfection of *D. melanogaster* clock, there is no detectable luciferase activity. The constitutively active form of CaMKII (mouse T286D) increases luciferase activity (normalized to CLOCK⁺; data are shown as mean \pm s.e.m.; $n = 4$ biological replicates). **c**, RNA sequencing

reads mapped to the *CaMKII.1* genomic locus. Arrows, major differences between the strains. **d**, Relative expression levels of the four major *CaMKII.1* transcripts (RA–RD) and the minor variant RO in the Por and Jean strains of *C. marinus*, as measured by qPCR (data are shown as mean \pm s.e.m.; two-sided Wilcoxon rank-sum test; *** $P < 0.0005$; * $P < 0.05$; NS, not significant; Holm correction for multiple testing; biological replicates, Por $n = 9$, Jean $n = 10$; except for RO: Por $n = 3$, Jean $n = 8$). RO was not detectable in six additional biological replicates of the Por strain, suggesting that the expression differences are even greater than currently estimated. Figure 3a shows the same data, normalized to the respective Por strain variants.

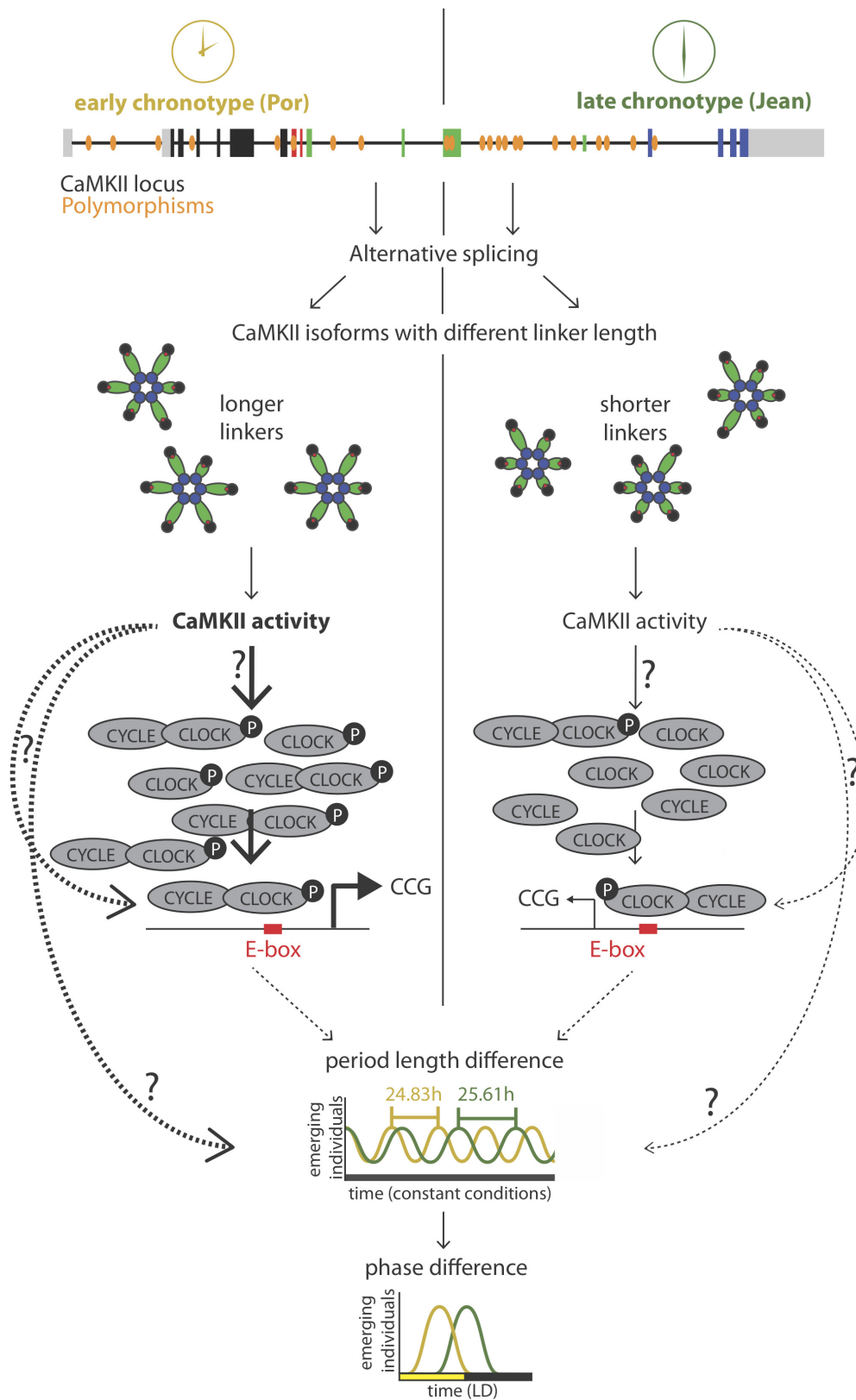
a

Por	AGTGTGGAAG	TAATTTTTAT	CTCACATCCA	TTAAGATTCT	TATTGCTTTT
Jean	AGTGTGGAAG	TAATTTTTAT	CTCACAGCCA	TTAAGATTCT	TATTGCTTTT
Por	GAATCCCTGA	AAGCTTCCCG	AGCGGCACAC	ACGGCAATGT	CTCTCACAAA
Jean	AAATCCCTGA	AAGCTT----	-----	-----	-----
Por	TCTAGCATAG	GTCTGCCATA	AGGCTAAAAA	AAATTGACAG	TTGCCGATAG
Jean	-----	-----	-----	-----	-----
Por	GTCATCCCCA	CGGTGCCACA	TGGTGCATTT	GTGCCGCTCG	GGCTTATACA
Jean	-----	-----	-----	-----	-ATACATACA
Por	TTGTACATTC	ATTGCACATT	ATAAATCCAT	AATTTTTTAAT	TTGATATCTA
Jean	TTGTACATTC	ATTGCACATT	ATAAATCCAT	AATTTTTTAAT	TTGATATCTA

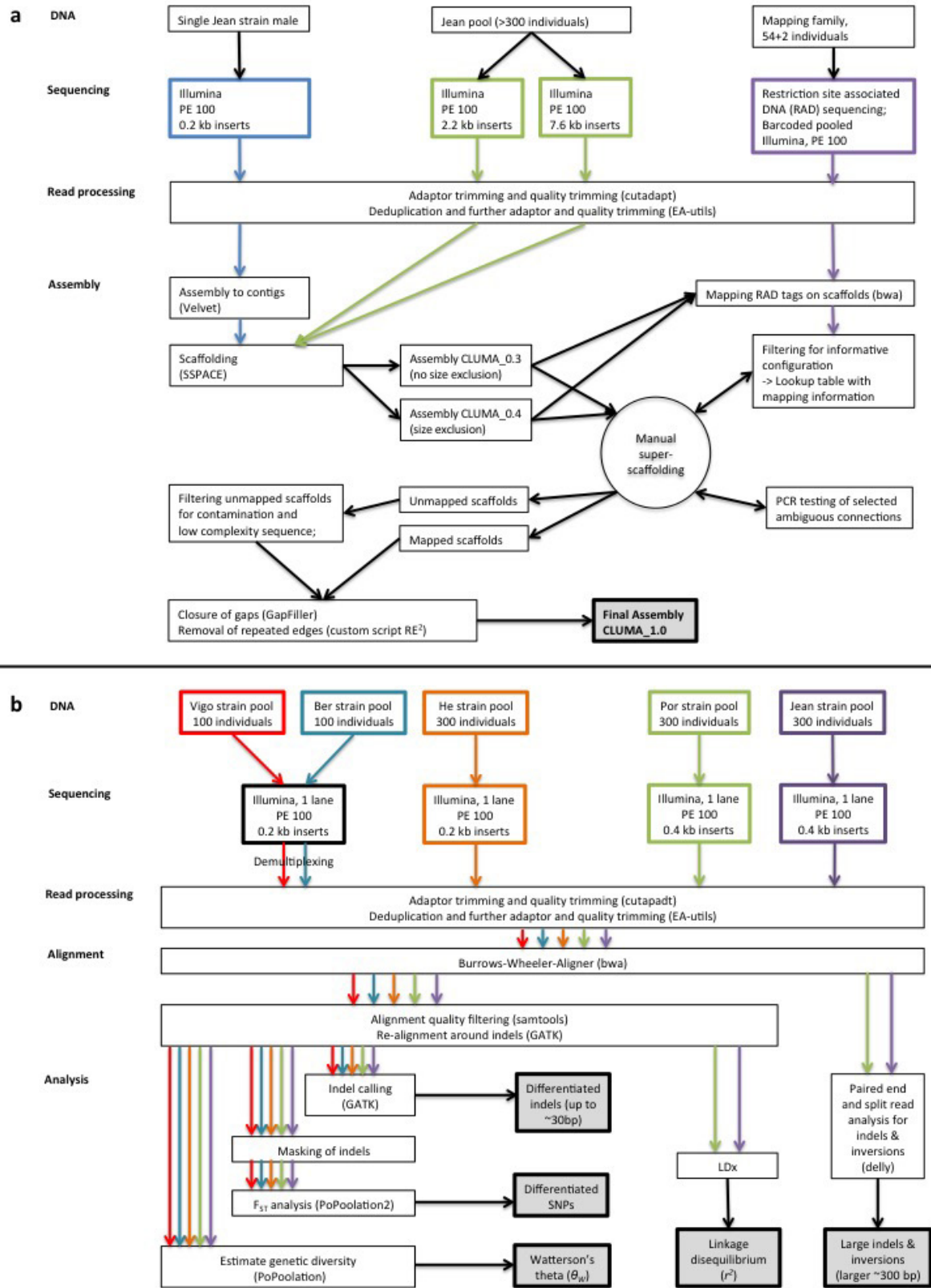


Extended Data Figure 7 | A differentiated 125-bp insertion in the CaMKII locus. **a**, Alignment of the part of the CaMKII locus of the Por and Jean strains that carries a 125-bp insertion in the Por strain. **b**, Pool-seq reads ($>150\times$ coverage) of this position for Por and Jean, as shown in the integrated genome viewer (IGV). The reference genome does not have the 125-bp insertion. At the position marked by the red box, the Jean strain has a 4-bp polymorphic indel (ATAC, frequently misaligned

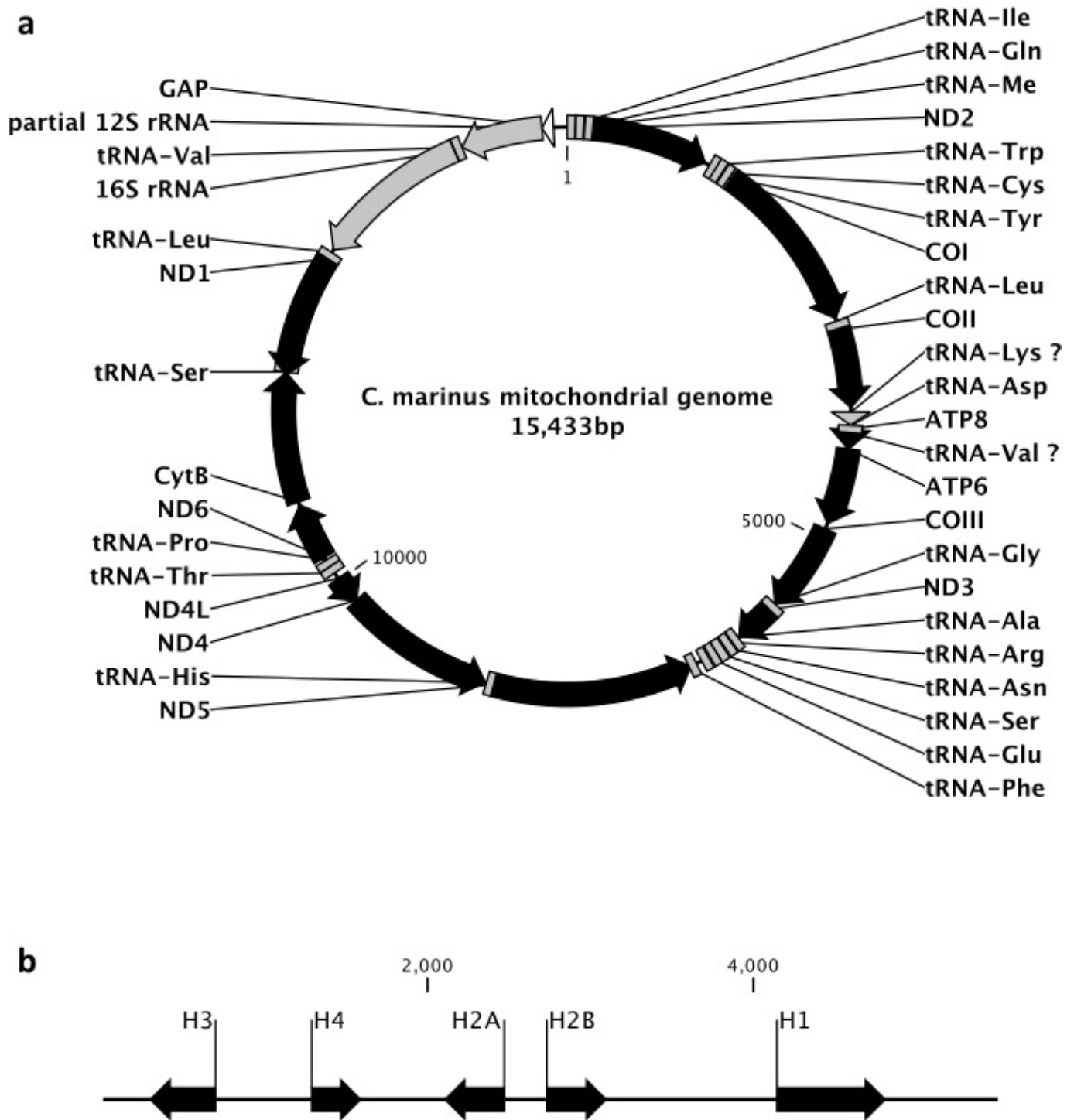
due to a SNP 8 bp downstream), whereas the Por strain has the 125-bp insertion (but not the 4-bp ATAC insertion). In Jean all reads span the indel, suggesting that if the 125-bp insertion is present in Jean at all, its frequency is very low. In contrast, in Por all reads but one end at this position, suggesting the frequency of the 125-bp insertion in Por is 154 of 155 reads or >0.99 .



Extended Data Figure 8 | Model of circadian timing adaptation via sequence differences in the *CaMKII.1* genomic locus. Exon coloration as in Fig. 2b. The arrows with question marks indicate possible pathways that alone or in combination could mediate the effect of *CaMKII.1* on timing. Dotted lines, indirect effects.



Extended Data Figure 9 | Analyses overview. **a**, Overview of the genome assembly process. **b**, Overview of the population genomic analyses.



Extended Data Figure 10 | Arrangement of the mitochondrial genome and of the histone gene cluster in *C. marinus*. **a, b,** mitochondrial genome (**a**) and histone gene cluster (**b**) arrangements in *C. marinus*. Protein-coding genes are shown in black, tRNAs and rRNAs in grey.

Correcting mitochondrial fusion by manipulating mitofusin conformations

Antonietta Franco^{1*}, Richard N. Kitsis^{2*}, Julie A. Fleischer¹, Evripidis Gavathiotis³, Opher S. Kornfeld⁴, Guohua Gong¹, Nikolaos Biris³, Ann Benz⁵, Nir Qvit⁴, Sara K. Donnelly⁶, Yun Chen², Steven Mennerick⁵, Louis Hodgson⁶, Daria Mochly-Rosen⁴ & Gerald W. Dorn II¹

Mitochondria are dynamic organelles that exchange contents and undergo remodelling during cyclic fusion and fission. Genetic mutations in *MFN2* (the gene encoding mitofusin 2) interrupt mitochondrial fusion and cause the untreatable neurodegenerative condition Charcot–Marie–Tooth disease type 2A (CMT2A). It has not yet been possible to directly modulate mitochondrial fusion, in part because the structural basis of mitofusin function is not completely understood. Here we show that mitofusins adopt either a fusion-constrained or a fusion-permissive molecular conformation, directed by specific intramolecular binding interactions, and demonstrate that mitofusin-dependent mitochondrial fusion can be regulated in mouse cells by targeting these conformational transitions. On the basis of this model, we engineered a cell-permeant minipeptide to destabilize the fusion-constrained conformation of mitofusin and promote the fusion-permissive conformation, reversing mitochondrial abnormalities in cultured fibroblasts and neurons that harbour CMT2A-associated genetic defects. The relationship between the conformational plasticity of mitofusin 2 and mitochondrial dynamism reveals a central mechanism that regulates mitochondrial fusion, the manipulation of which can correct mitochondrial pathology triggered by defective or imbalanced mitochondrial dynamics.

Mitochondria fuse together to transfer genetic information and promote mutual repair through content exchange. Tethering and fusion of the mitochondrial outer membrane is mediated by mitofusins (Mfn) 1 and 2 and is essential both for embryonic development^{1–3} and tissue homeostasis^{4–6}. The multiple mutations that provoke MFN2 dysfunction can cause the untreatable neurodegenerative condition CMT2A⁷. The obligate initial step in mitochondrial fusion is formation of mitochondria–mitochondria tethers. Mitofusins on different mitochondria undergo GTPase-independent trans-dimerization through the anti-parallel binding of their extended C-terminal α -helical domains⁸. Mitofusins are conventionally thought to exist in a constitutively ‘active’ extended molecular conformation that is permissive of tethering; however, there is a lack of specific contextual data on other possible structural determinants of mitofusin-mediated mitochondrial tethering. Here, on the basis of domain similarity between human mitofusins and the prototypical dynamin superfamily protein bacterial dynamin-like protein (DLP)⁹, we posit that mitofusins can adopt a more folded molecular conformation. Such conformational plasticity could be experimentally and therapeutically manipulated to enhance or suppress mitochondrial tethering and subsequent tethering-dependent organelle fusion.

Conformational malleability of MFN2

MFN1 and MFN2 share an N-terminal GTPase domain followed by a first coiled-coil heptad-repeat region (HR1), two adjacent small transmembrane domains and a second, C-terminal, coiled-coil heptad-repeat region (HR2) (Fig. 1a). Amino-acid conservation is highest in the GTPase, transmembrane and HR2 regions (Extended Data Fig. 1a). HR2 domains that extend from MFN molecules on different

mitochondria dimerize to tether the organelles⁸ (Fig. 1b) and HR2 can also bind HR1 (ref. 10). Using the crystal structure of bacterial DLP⁹ to model the structure of MFN2, we refined the identities of HR2 amino acids that mediate intermolecular HR2–HR2 tethering⁸ (Extended Data Fig. 1b and Supplementary Video 1). We also predicted that these same amino acids mediate the intramolecular antiparallel binding of HR2 to HR1 (Fig. 1c, Extended Data Fig. 1c and Supplementary Video 2). Local unfolding and bending of the HR1 α -helix at amino acids 384–387 (Arg–Glu–Gln–Gln) constrains the HR2 binding region to within the core MFN2 globular structure. This restrains HR2, which we predict to bend at amino acids 712–715 (Gln–Glu–Ile–Ala) (Figs 1c, d and Supplementary Video 3). This configuration would not be conducive for mitochondrial tethering, which requires the cytosolic extension of HR2 (Fig. 1d and Supplementary Video 4).

Manipulating fusion with HR1 minipeptides

We examined the functional importance of intramolecular binding between MFN2 HR1 and HR2 by introducing a recombinant MFN2 HR1 minipeptide fragment on an adenovirus vector, containing the sequences predicted to mediate intramolecular HR1–HR2 binding (MP-HR1) into mouse embryonic fibroblasts (MEFs). According to our proposed mechanism, the competition for intramolecular HR1–HR2 interactions by free HR1 domains should unfold the protein, extend the terminal HR2 arm and promote mitochondrial tethering or fusion. Indeed, MP-HR1 increased the mitochondrial aspect ratio (the ratio of the mitochondrial major axis dimension (or length) to minor axis dimension (or width), reflecting fusion) in both wild-type and Mfn2-null¹¹ MEFs (Fig. 1e). Thus, conformational plasticity and

¹Center for Pharmacogenomics, Department of Internal Medicine, Washington University School of Medicine, St. Louis, Missouri, USA. ²Departments of Medicine (Cardiology) and Cell Biology and Wilf Family Cardiovascular Research Institute, Albert Einstein College of Medicine, Bronx, New York, USA. ³Departments of Biochemistry and Medicine, Wilf Family Cardiovascular Research Institute, Albert Einstein Cancer Center, Albert Einstein College of Medicine, Bronx, New York, USA. ⁴Department of Chemical and Systems Biology, Stanford University, School of Medicine, Stanford, California, USA. ⁵Department of Psychiatry and Taylor Family Institute for Innovative Psychiatric Research, Washington University School of Medicine, St. Louis, Missouri, USA. ⁶Department of Anatomy and Structural Biology, Gruss-Lipper Biophotonics Center, Albert Einstein College of Medicine, Bronx, New York, USA.

*These authors contributed equally to this work.

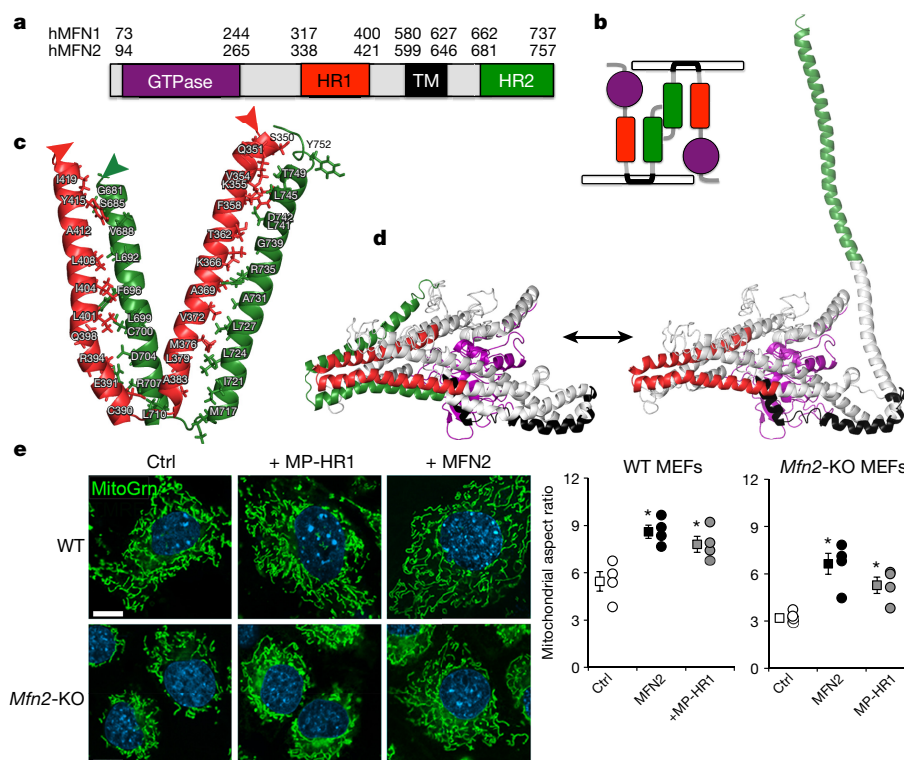


Figure 1 | Intramolecular interactions between MFN2 HR1 and HR2 modulate mitochondrial fusion. **a**, Domain model of human MFN structure. The GTPase domain is purple, HR1 is red, transmembrane (TM) is black, HR2 is green. **b**, Trans-MFN-MFN interaction, producing the mitochondrial tethering that precedes outer membrane fusion. **c**, Intramolecular human MFN2 HR1 (red)-HR2 (green) interactions based on crystal structures of bacterial DLP. **d**, Homology model of human MFN2 showing putative constrained/inactive (left) and extended/active (right)

conformations. **e**, Confocal images of wild-type (WT) or *Mfn2*-knockout (KO) MEFs expressing β -gal (Ctrl), human MFN2 MP-HR1 minipeptide fragment or fully functional wild-type human MFN2. MitoTracker Green (MitoGrn) stains mitochondria, blue Hoechst stains nuclei. Quantitative results are shown on the right. Group data are mean \pm s.e.m.; * $P < 0.05$ versus control (ANOVA) in four independent experiments for each condition. Scale bars, 10 μ m.

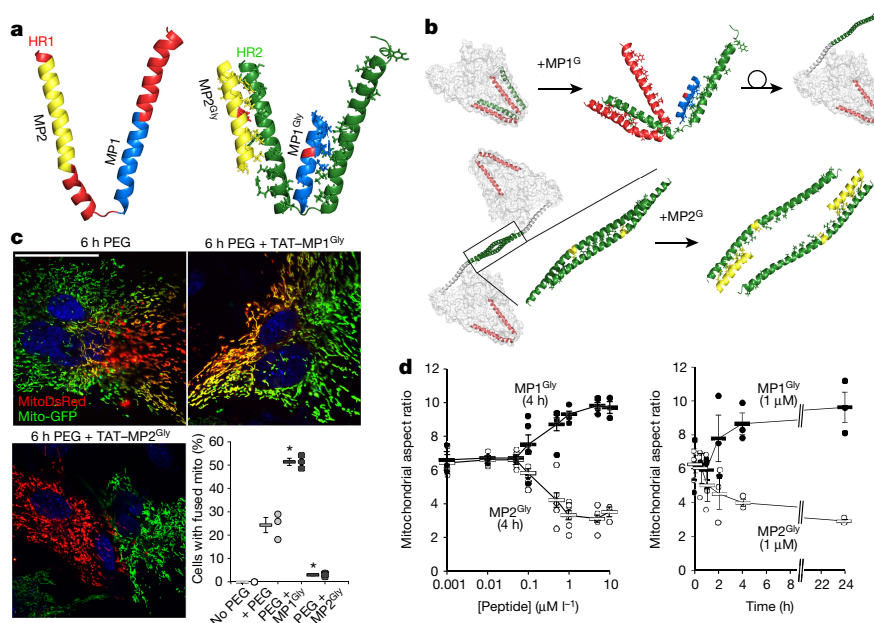


Figure 2 | Adenovirally expressed Glycyl-substituted minipeptides derived from MFN2 HR1 regulate Mfn-mediated mitochondrial fusion.

a, Derivation of HR1 minipeptides and their predicted binding to HR2. On the right panel, red shows location of the critical glycine substitution. **b**, Ribbon representations showing (top) TAT-MP1^{Gly} (blue) interrupting intramolecular HR1-HR2 binding, and (bottom) TAT-MP2^{Gly} (yellow) interrupting the intermolecular HR2-HR2 interaction essential to

mitochondrial tethering. **c**, Mitochondrial content exchange (red-green overlay) 6 h after polyethylene glycol (PEG)-mediated cell fusion. Scale bar, 20 μ m. $n = 3$ experiments measuring ~ 50 cells each; * $P < 0.01$ versus those treated with PEG alone (ANOVA and Tukey's test). **d**, TAT-mini-peptide dose-response at 4 h (left, $n = 6$ or 7) and time-course of 1 μ M effects (right, $n = 3$). Group data are mean \pm s.e.m.

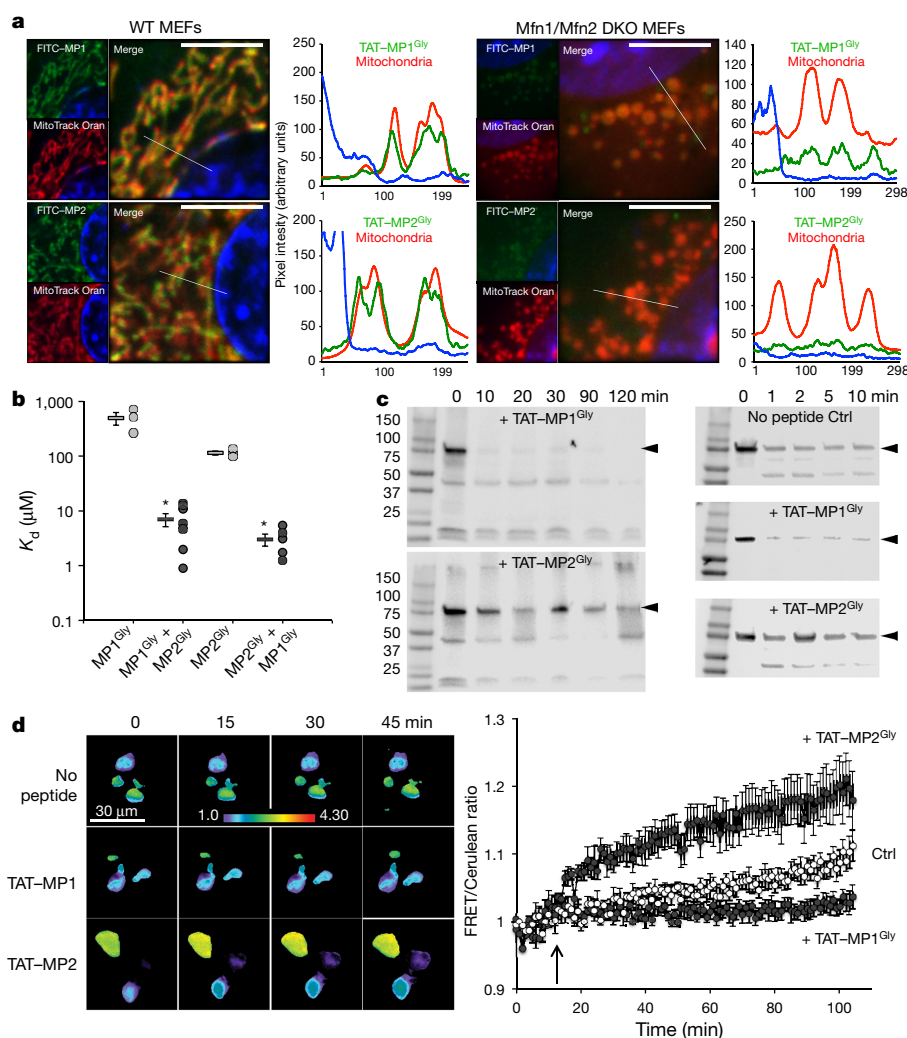


Figure 3 | Molecular effects of HR1-derived TAT-mini-peptides on MFN2. **a**, Confocal localization of FITC-labelled TAT-mini-peptides (green) in wild-type (WT) MEFs (left) and Mfn null MEFs (*Mfn1/Mfn2* double knockout (DKO); right). Red mitochondria were stained with MitoTracker Orange; blue is nucleus (Hoechst). Line-scans are to the right of their respective images. Representative of >20 images per group. Scale bars, 5 μm. **b**, Peptide binding to Mfn2. Group data are mean ± s.e.m.;

$n = 3$ studies for baseline and 6 or 7 studies for cooperativity. * $P < 0.05$ versus same peptide alone (t -test). **c**, Carboxypeptidase protection assays. Anti-Mfn2 recognizes HR2 (human MFN2 amino acids 661–758). Arrows show MFN2; gel source images are shown in Extended Data Fig. 1. **d**, MFN2 live cell FRET studies. Representative images are on the left; quantitative data (mean ± s.e.m., $n = 10$) are to the right. Arrow indicates time of TAT-peptide addition.

reversible intramolecular HR1–HR2 binding can be as important as intermolecular HR2–HR2 binding for Mfn function.

To localize HR1–HR2 binding domains, we used an adenovirus vector to express mini-peptides that flank the HR1 Arg–Glu–Gln–Gln bend. These were mini-peptide 1 (MP1, a mini-peptide made up of residues 367–384 of MFN2), MP2 (made up of residues 398–418) and MP3 (made up of residues 428–448). We also engineered mutated mini-peptides, substituting the centrally positioned leucine residues for either glycine residues (to increase flexibility in the secondary structure) or proline residues (to break the α -helix). MP3 and its Gly- and Pro-substituted mutant counterparts did not affect mitochondrial aspect ratio, whereas MP1^{Gly} and MP2^{Gly} (the glycine-substituted mini-peptide counterparts of MP1 and MP2) consistently had opposing effects on mitochondrial morphology (Extended Data Fig. 2).

MP-HR1, MP1^{Gly} and MP2^{Gly} all altered the mitochondrial aspect ratio in MEFs that expressed Mfn1 or Mfn2 in any combination, but not in MEFs that completely lack mitofusins (Extended Data Fig. 2a). Thus, MFN2 HR1-derived mini-peptides modulate mitochondrial fusion, mediated by either Mfn1 or Mfn2. MP1^{Gly} and MP2^{Gly} did not adversely affect mitochondrial polarization status (Extended Data Fig. 2b), or Parkin recruitment and mitophagy induced by

mitochondrial uncoupling (Extended Data Fig. 2c), which are also affected by mitofusins^{12,13}.

According to our molecular model of MFN2, the fusion-promoting MP1^{Gly} should bind amino acids 716–736, near to the MFN2 C terminus (Fig. 2a). Competition for intramolecular HR1–HR2 binding by the peptide would release HR2 from its HR1 anchor and permit it to unfold into a cytosol-extended conformation (Fig. 2b). By contrast, the fusion-suppressing MP2^{Gly} should bind HR2 at amino acids 682–701 (Fig. 2a), thus interrupting the HR2–HR2 interactions that are essential for transmembrane tethering (Fig. 2b).

We studied our mini-peptides after synthesizing them as cell-permeant C-terminal TAT_{47–57} (residues 47–57 of the HIV transactivator protein, which are able to directly transport through the cell membrane) conjugates. The fusion-mediated mixing of mitochondrial contents¹⁴ was more than doubled by the addition of 1 μM TAT-MP1^{Gly} for 6 h, whereas 1 μM TAT-MP2^{Gly} completely suppressed mitochondrial fusion (Fig. 2c). The mitochondrial aspect ratio changed in parallel (Fig. 2d), with EC₅₀ values of 280 ± 90 nM and 250 ± 90 nM for TAT-MP1^{Gly} and TAT-MP2^{Gly} effects, respectively (Fig. 2d and Extended Data Fig. 3a). The abundance of proteins involved in mitochondrial dynamics (Extended Data Fig. 3b, c), MFN2 GTPase activity

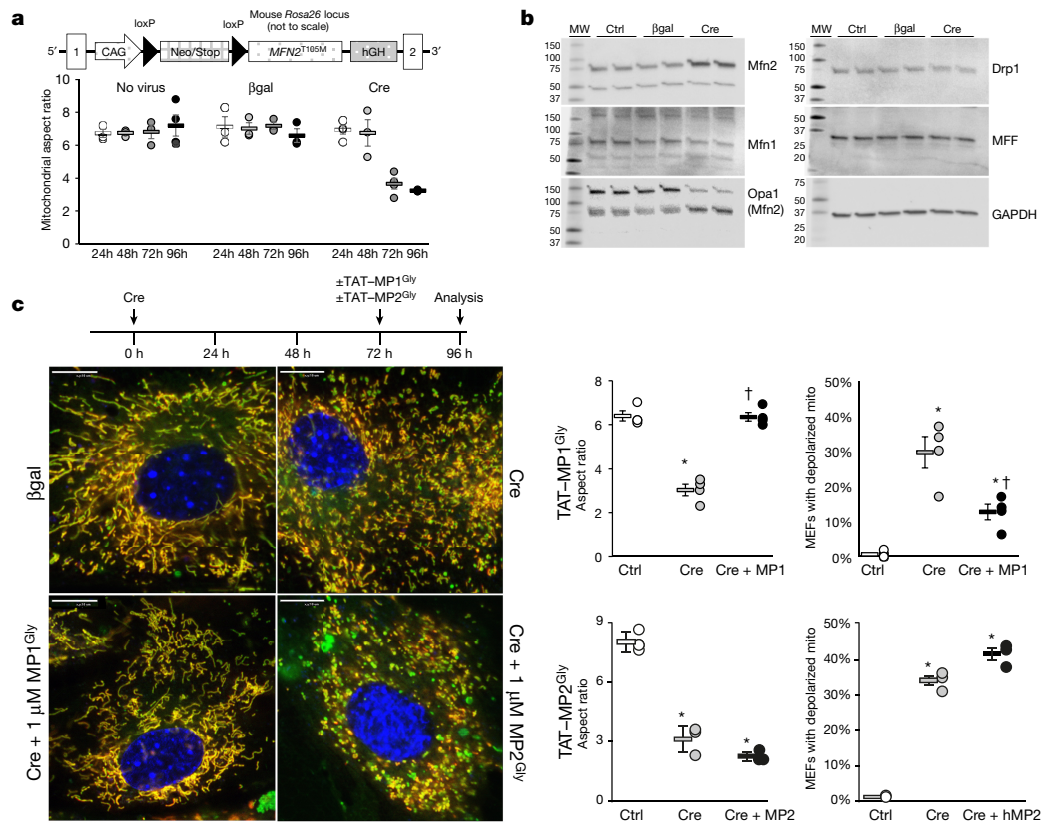


Figure 4 | TAT-MP1^{Gly} normalizes mitochondrial dysmorphology induced by the CMT2A mutant *MFN2*^{T105M}. **a**, Schematic representation of the conditional *MFN2*^{T105M} transgene construct (top). Time course of mitochondrial fragmentation after Cre-mediated induction of *MFN2*^{T105M} (bottom). Group data are mean ± s.e.m., *n* = 3; **P* < 0.05 versus 24 h same treatment group (ANOVA). **b**, Immunoblot analysis of mitochondrial dynamics factors in *MFN2*^{T105M} MEFs treated as indicated above; gel source data are in Extended Data Fig. 1 and quantitation is in Extended Data Fig. 8d. **c**, Confocal micrographs (representative of >15 per group) showing mitochondrial fragmentation evoked by *MFN2*^{T105M}, its normalization 24 h after application of 1 μM TAT-MP1^{Gly}, and its exaggeration 24 h after 1 μM TAT-MP1^{Gly}. Experimental design is depicted above. Scale bars, 10 μm. On the right are group data; *n* = 3 or 4. Group data are mean ± s.e.m.; **P* < 0.05 versus control, †*P* < 0.05 versus Cre (ANOVA).

Fig. 8d. MW, molecular weight. **c**, Confocal micrographs (representative of >15 per group) showing mitochondrial fragmentation evoked by *MFN2*^{T105M}, its normalization 24 h after application of 1 μM TAT-MP1^{Gly}, and its exaggeration 24 h after 1 μM TAT-MP1^{Gly}. Experimental design is depicted above. Scale bars, 10 μm. On the right are group data; *n* = 3 or 4. Group data are mean ± s.e.m.; **P* < 0.05 versus control, †*P* < 0.05 versus Cre (ANOVA).

(Extended Data Fig. 4a), and mitochondrial polarization and cell viability (Extended Data Fig. 4b) were unaffected by the minipeptides.

HR1 minipeptides alter MFN2 conformation

Additional studies validated the conformational-shift mechanism of minipeptide action: FITC-labelled TAT-conjugated minipeptides decorated mitochondria in normal cells, but not in cells that lacked mitofusins (Mfn1/Mfn2 double-knockout MEFs) (Fig. 3a). The binding of minipeptides to recombinant Mfn2 was positively cooperative (Fig. 3b), consistent with the destabilization of highly folded Mfn2 by MP1^{Gly} and thus facilitating the binding of MP2^{Gly} to newly extended C-terminal HR2 domains. We demonstrated that TAT-MP1^{Gly} promoted MFN2 HR2 unfolding and extension by showing that it accelerated C-terminal-directed proteolytic digestion of MFN2 (TAT-MP2^{Gly} had the opposite effect) (Fig. 3c). Finally, using Förster resonance energy transfer (FRET), we determined that TAT-MP2^{Gly} promotes a more folded conformation of human MFN2, tagged at the N terminus with mCerulean1 and the C terminus with mVenus1 (Fig. 3d). Collectively, the data in Figs. 2, 3 demonstrate that the effects of MFN2 HR1-derived minipeptides on mitochondrial fusion derive from their ability to promote or suppress the unfolding of Mfn2 and the extension of its HR2 domain.

Correcting mitochondrial fusion defects

Pathological mitochondrial fragmentation in CMT2A is caused by nonsense or damaging non-synonymous *MFN2* mutations⁷. We simulated compound nonsense *MFN2* mutations seen in autosomal recessive CMT2A by *Mfn2* gene ablation. TAT-MP1^{Gly} treatment for

24 h normalized the mitochondrial aspect ratio of Mfn2-knockout MEFs (Extended Data Fig. 5). However, after the adenoviral introduction of mouse *Mfn2*^{R109A} (a GTPase-defective mutant), TAT-MP1^{Gly} did not correct mitochondrial fragmentation in Mfn1/Mfn2-null MEFs¹⁵ (although mitochondrial clumping suggested enhanced mitochondria-mitochondria tethering in the absence of organelle fusion) (Extended Data Fig. 6). An Mfn2 amino-acid mutant with normal fusion activity but defective Parkin binding (T111 and S442 alanine mutants^{12,13}) responded similarly to wild-type Mfn2.

These findings hint at the therapeutic potential of compelling endogenous normal mitofusins into their active, unfolded conformation in order to overcome dominant inhibition by GTPase-defective MFN2 mutants in diseases like CMT2A. To test this notion, we used MEFs that express the CMT2A-causative mutant human *MFN2*^{T105M} (refs 16,17) as a conditional *flax/stop* (*fl/st*) transgene (Figs 4a, b). Mitochondria of *MFN2*^{T105M} *fl/st* MEFs fragmented and depolarized after adenoviral-Cre application (Figs 4a, c), but addition of 1 μM TAT-MP1^{Gly} rapidly reversed mitochondrial abnormalities. By contrast, suppressing mitochondrial fusion with TAT-MP2^{Gly} aggravated mitochondrial dysmorphology and exacerbated the mitochondrial depolarization that was induced by *MFN2*^{T105M} (Fig. 4c). Parallel adenoviral expression of wild-type MFN2 at more than twice the levels of *MFN2*^{T105M} (fivefold versus twofold) provoked neither mitochondrial nor cellular toxicity (Extended Data Fig. 7).

Repairing neuronal CMT2A pathology

As neurons are the cell type most relevant to CMT2A, we examined the effects of TAT-MP1^{Gly} on cultured rat neurons bearing the

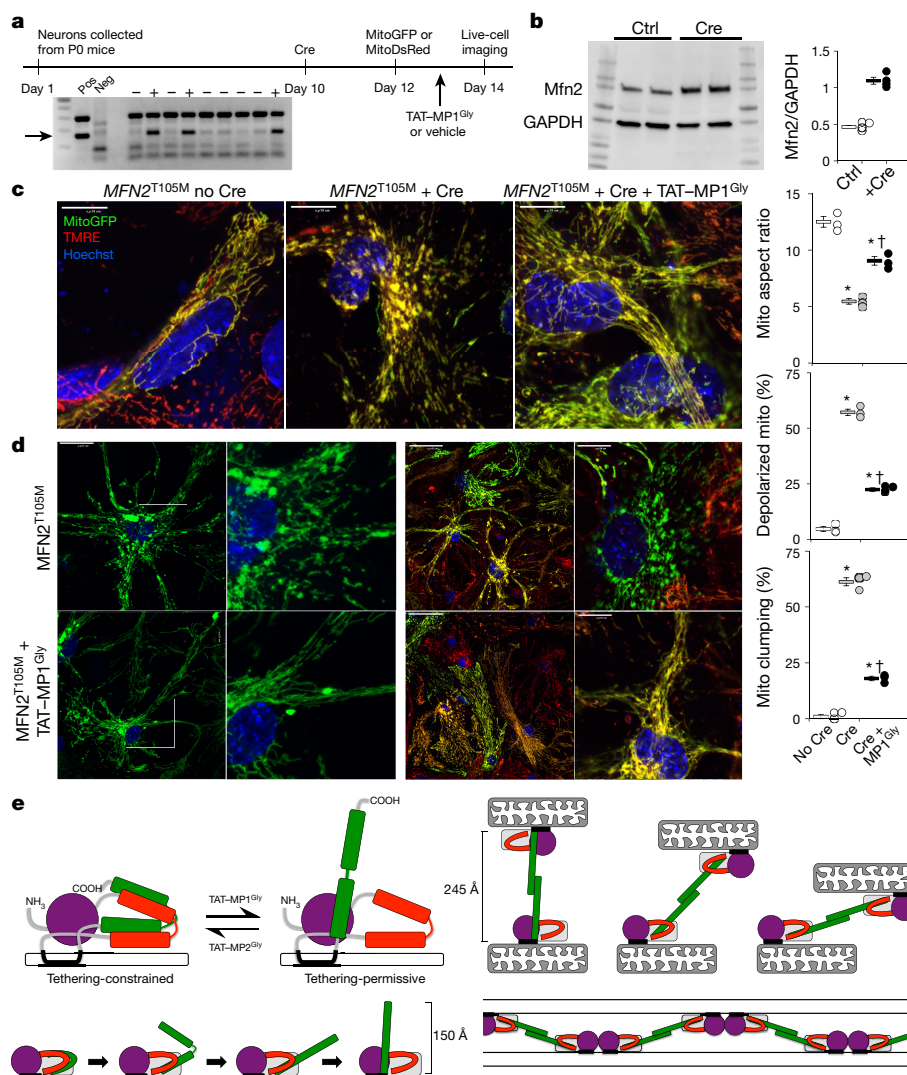


Figure 5 | TAT-MP1^{Gly} corrects mitochondrial pathology in CMT2A *MFN2*^{T105M} mouse neurons. **a**, Experimental design. Inset shows PCR genotyping of a *MFN2*^{T105M} *fl/st* transgene (arrow) mouse litter whose pups were used for neuron harvesting and culture. **b**, Immunoblot of Mfn2 expression in *MFN2*^{T105M} *fl/st* cortical neurons 72 h after adenoviral transfection of Cre or β -gal (control) (quantitative data to the right are $n = 4$); gel source images are shown in Extended Data Fig. 1. **c**, Merged live confocal images of cultured neurons without (no Cre) and with (with Cre) *MFN2*^{T105M} induction, and in *MFN2*^{T105M}-induced neurons 24 h after addition of 1 μ M TAT-MP1^{Gly}. Scale bars, 10 μ m. **d**, *MFN2*^{T105M} expressing neurons without (top) and with (bottom) 24-h treatment with

1 μ M TAT-MP1^{Gly}. Scale bars, 40 μ m (left two images of each panel); 10 μ m (right two images of each panel). Quantitative data are to the right; each point is a result from one of three independently-established *MFN2*^{T105M} neuronal cultures, each with three biological replicates of ~ 20 neurons. Group data are mean \pm s.e.m.; * $P < 0.01$ versus no Cre; † $P < 0.01$ versus *MFN2*^{T105M} + Cre. **e**, (left) Schematic depiction of the folded (with HR2 constrained; left) and unfolded (with HR2 extended; right) mitofusin conformations; HR2 unfolding is below. HR2 flexing at Gly 623/642 after trans-association of HR2 domains (right); molecular patch formed between mitochondria by coincident GTP-dependent/GTPase domain-mediated *cis* Mfn2 dimerization is below.

GTPase-defective *MFN2*^{K109A} mutant and found that its addition corrected mitochondrial pathology (Extended Data Fig. 8). We tested fusion-promotion further through the conformational manipulation of mitofusins in cultured neurons from mouse pups carrying the conditional *MFN2*^{T105M} *fl/st* expression allele, a 'disease in a dish' model of a naturally occurring human CMT2A mutation (Fig. 5a). Live-cell analysis demonstrated that adeno-Cre induction of *MFN2*^{T105M} (Fig. 5b) evoked widespread neuronal mitochondrial dysmorphology, with fragmentation and partial depolarization (Fig. 5c, d and Extended Data Fig. 9). Mitochondrial clumping was prevalent after induction of *MFN2*^{T105M}, which was consistent with the interruption of GTP-dependent mitochondrial fusion but preserved GTPase-independent mitochondrial tethering (Fig. 5d). Addition of 1 μ M TAT-MP1^{Gly} for 24 h largely reversed these abnormalities (Fig. 5c, d and Extended Data Fig. 9).

Discussion

By combining computational modelling of the structure of mitofusins (based on the crystal structure of bacterial DLP) with functional interrogation of intramolecular mitofusin interactions using engineered competing peptides, we provide evidence for two functionally distinct conformational states of mammalian mitofusins. The contrasting effects of the engineered minipeptides on mitochondrial fusion were linked to their reciprocal modulation of mitofusin HR1 folding and unfolding.

The canonical representation of MFN1 and MFN2, as described by Chan and colleagues⁸, is of a protein anchored to the outer mitochondrial membrane by a transmembrane domain, with N- and C-terminal structures extending perpendicularly into the cytosol. This resembles an unfolded MFN conformation that is optimal for mitochondrial tethering and therefore permissive of fusion. Our model suggests,

however, that the core globular MFN molecule is adherent to the mitochondrial membrane. Moreover, in the resting state, or that non-permissive to tethering, the HR2 domain is restrained by antiparallel intramolecular binding to HR1; destabilization of this binding allows HR2 to unfold and extend into the cytosol, into a tethering-permissive state.

Conformational plasticity has important implications for the mitochondrial gap-distance that is maintained by Mfn–Mfn tethering, variously reported as 78 ± 37 Å (ref. 18) or 159 ± 30 Å (ref. 8). On the basis of a calculated HR2 arm-length of 150 Å that begins from the putative Gly ‘shoulder’ at the C terminus of the transmembrane domain (Gly623 of MFN1; Gly642 of MFN2), and accounting for the 60-amino-acid overlap of the HR2–HR2 homodimer, the maximal tethered mitochondrial gap distance according to our model would be 245 Å. However, flexing of each MFN HR2 at its shoulder would retract tethered mitochondria (Fig. 5e), narrowing the gap between them. The crystal structure of bacterial DLP indicates that the binding of GTP promotes DLP dimerization at GTP-binding domains⁹. The same GTP-dependent event for MFN2 would, with concomitant trans-dimerization of HR2 domains, create a multimeric molecular patch between two adjoining mitochondria, facilitating GTP-dependent outer-membrane fusion (Fig. 5e). Thus, MFN structural malleability may be important not only to initiate tethering, but to facilitate the progression from organelle tethering to apposition and then union.

In designing our conformation-altering minipeptides, we contemplated the difficulty of destabilizing a constrained MFN structure using peptides identical to the parent HR1 domains with which they were intended to compete. Accordingly, we substituted rotational Gly residues for Leu residues facing away from the putative HR1–HR2 (MP1^{Gly}) and HR2–HR2 (MP2^{Gly}) interaction sites. The importance of minipeptide flexibility at these sites was confirmed by their functional inactivation by substitution with rigid proline residues at the same positions.

Our fusogenic cell-permeant peptide, TAT–MP1^{Gly}, reversed mitochondrial dysmorphology and depolarization both in otherwise normal MEFs and in cultured neurons expressing either the artificial GTPase-deficient Mfn2^{K109A} mutant or the naturally occurring human CMT2A GTPase mutant, Mfn2^{T105M}. Because TAT–MP1^{Gly} did not correct the mitochondrial pathology induced by Mfn2^{K109A} in mitofusin-null cells, we conclude that rescue of the *in vitro* CMT2A models is due to enhanced mitochondrial tethering and fusion mediated by endogenous normal Mfn1 and Mfn2. Most CMT2A mutations, including Mfn2^{T105M}, are autosomal dominant—patients have one mutant *MFN2* allele, one normal *MFN2* allele, and two normal *MFN1* alleles⁷. Thus, rather than relying exclusively on genetic engineering to correct or silence *MFN2* gene mutations, pharmacologically unfolding endogenous normal mitofusins to promote mitochondrial tethering and enhance fusion could prove beneficial in CMT2A and pathophysiologically related diseases.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 May; accepted 14 October 2016.

Published online 24 October 2016.

1. Chan, D. C. Fusion and fission: interlinked processes critical for mitochondrial health. *Annu. Rev. Genet.* **46**, 265–287 (2012).

2. Chen, H. *et al.* Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. *J. Cell Biol.* **160**, 189–200 (2003).
3. Kasahara, A., Cipolat, S., Chen, Y., Dorn, G. W., II & Scorrano, L. Mitochondrial fusion directs cardiomyocyte differentiation via calcineurin and Notch signaling. *Science* **342**, 734–737 (2013).
4. Chen, H., McCaffery, J. M. & Chan, D. C. Mitochondrial fusion protects against neurodegeneration in the cerebellum. *Cell* **130**, 548–562 (2007).
5. Chen, H. *et al.* Mitochondrial fusion is required for mtDNA stability in skeletal muscle and tolerance of mtDNA mutations. *Cell* **141**, 280–289 (2010).
6. Song, M., Mihara, K., Chen, Y., Scorrano, L. & Dorn, G. W., II. Mitochondrial fission and fusion factors reciprocally orchestrate mitophagic culling in mouse hearts and cultured fibroblasts. *Cell Metab.* **21**, 273–285 (2015).
7. Bombelli, F. *et al.* Charcot-Marie-Tooth disease type 2A: from typical to rare phenotypic and genotypic features. *JAMA Neurol.* **71**, 1036–1042 (2014).
8. Koshihara, T. *et al.* Structural basis of mitochondrial tethering by mitofusin complexes. *Science* **305**, 858–862 (2004).
9. Low, H. H. & Löwe, J. A bacterial dynamin-like protein. *Nature* **444**, 766–769 (2006).
10. Huang, P., Galloway, C. A. & Yoon, Y. Control of mitochondrial morphology through differential interactions of mitochondrial fusion and fission proteins. *PLoS One* **6**, e20655 (2011).
11. Chen, H. *et al.* Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. *J. Cell Biol.* **160**, 189–200 (2003).
12. Chen, Y. & Dorn, G. W., II. PINK1-phosphorylated mitofusin 2 is a Parkin receptor for culling damaged mitochondria. *Science* **340**, 471–475 (2013).
13. Gong, G. *et al.* Parkin-mediated mitophagy directs perinatal cardiac metabolic maturation in mice. *Science* **350**, aad2459 (2015).
14. Chen, H., Chomyn, A. & Chan, D. C. Disruption of fusion results in mitochondrial heterogeneity and dysfunction. *J. Biol. Chem.* **280**, 26185–26192, (2005).
15. Detmer, S. A. & Chan, D. C. Complementation between mouse Mfn1 and Mfn2 protects mitochondrial fusion defects caused by CMT2A disease mutations. *J. Cell Biol.* **176**, 405–414 (2007).
16. Calvo, J. *et al.* Genotype-phenotype correlations in Charcot-Marie-Tooth disease type 2 caused by mitofusin 2 mutations. *Arch. Neurol.* **66**, 1511–1516 (2009).
17. Detmer, S. A., Vande Velde, C., Cleveland, D. W. & Chan, D. C. Hindlimb gait defects due to motor axon loss and reduced distal muscles in a transgenic mouse model of Charcot-Marie-Tooth type 2A. *Hum. Mol. Genet.* **17**, 367–375 (2008).
18. Picard, M. *et al.* Trans-mitochondrial coordination of cristae at regulated membrane junctions. *Nat. Commun.* **6**, 6259 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements Supported by National Institutes of Health grants HL59888, HL128441 (G.W.D.); HL128071 (R.N.K. and G.W.D.); CA178394 (E.G.); MH078823 (S.M.); CA205262 (L.H.); and HL52141 (D.M.-R.); and American Cancer Society grant PF-15-135-01-CSM (S.K.D.). A.F. is supported by a fellowship from Società Italiana Ipertensione Arteriosa (SIIA). G.W.D. is the Philip and Sima K. Needleman-endowed Professor.

Author Contributions G.W.D., D.M.-R., R.N.K., L.H., and S.M. conceived or designed the research. G.W.D. wrote the manuscript. A.F., J.A.F., G.G. performed functional studies of minipeptides in MEFs. A.F., A.B., O.K. performed functional studies of minipeptides in neurons. A.F. and G.W.D. performed minipeptide mitochondrial colocalization. A.F. performed immunoblotting. N.Q. performed peptide-Mfn2 binding studies. O.K. performed Mfn2 GTPase assays. S.K.D., L.H., Y.C. performed Mfn2 FRET. E.G. and N.B. performed computational modelling and domain-interaction analyses. E.G., N.B., G.W.D., R.K. generated structural images.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.W.D. (gdorn@wustl.edu).

Reviewer Information Nature thanks R. Baloh, R. Lightowers and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Cell lines. Wild-type MEFs were prepared from embryonic age (E)10.5 C57BL/6 mouse embryos. SV-40 T-antigen-immortalized Mfn1-null (CRL-2992), Mfn2-null (CRL-2993) and Mfn1/Mfn2-double-null MEFs (CRL-2994), originally developed by D. Chan², were purchased from ATCC. MEFs conditionally expressing the human Mfn2^{T105M} mutation were prepared from E10.5 embryos of C57BL/6-*Gt(ROSA)26Sor[tm1(CAG-MFN2*^{T105M})Dple]/J* mice purchased from The Jackson Laboratory (Stock No: 025322). All cell lines were tested for mycoplasma and found to be mycoplasma-free. MEFs were maintained in DMEM containing 4.5 g l⁻¹ glucose (or galactose, where specified) supplemented with 10% fetal bovine serum, 1 × nonessential amino acids, 2 mM L-glutamine, 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin.

Cultured mouse neurons. Hippocampal and cortical neurons were prepared from individual postnatal day (P)0–1 mouse pups using a modification of methods described for rat neuronal culture¹⁹. Mouse pups were genotyped after neuron preparation to determine the presence or absence of the *MFN2*^{T105M} transgene; neurons derived from mice lacking the transgene were designated as normal controls. After 10 days of culture, neurons were infected with Adeno-Cre or Adeno-βgal control (50 MOI). After 72 h, 1 µM TAT-MP1^{Gly} or vehicle was added. Live neurons were imaged as described below.

Cultured rat motor neurons. Motor neurons were isolated from E17 Sprague-Dawley rats and cultured as described²⁰. After 10 days in culture, neurons were infected twice, 8 h apart, with 50 MOI of adenoviral wild-type Mfn2 or Mfn2^{K109A} and treated with 1 µM TAT-MP1^{Gly} or control TAT peptide for 24 h. The mitochondria of formalin-fixed neurons were labelled with anti-TOM20 (Santa Cruz Biotechnology); neurons were stained with Hoechst 33342. Laser confocal imaging using a 63× oil objective was used to identify isolated mitochondria to determine aspect ratio.

Adenoviral stocks. Empty adenovirus (Ad-CMV-Null; #1300) and adenovirus expressing β-galactosidase (Ad-CMV-b-Gal; #1080) or Cre recombinase (Ad-CMV-iCre; #1045) were obtained from Vector Biolabs. Adenoviral expression vectors for human Mfn2 HR1 and the minipeptides and their Gly- or Pro-mutant counterparts were generated and amplified using standard methods. Adenoviral vectors at an MOI of 100 were added to MEFs when at 50% confluence. Adeno-PA-GFP (mitoGFP) and adeno-mitoDsRed2 (both from SignaGen Laboratories) were added to cultured mouse neurons at an MOI of 10 or 50.

Antibodies and immunoblotting. Anti-Mfn1 (sc-50330) and anti-Opal (sc393296) were obtained from Santa Cruz. Anti-Mfn2 (ab56889), anti-Drp1 (ab56788), and anti-GAPDH (ab8245) were obtained from Abcam. Anti-MFF (17090-1-ap) was obtained from Proteintech. Cell-protein lysates were size-separated on 10% SDS-PAGE gels, transferred onto nylon membranes, incubated with primary antibody and detected with peroxidase-conjugated anti-mouse secondary antibody (Cell Signaling, 7076S, 1:4,000) or goat anti-rabbit secondary antibody (Thermo Scientific, 31460, 1:4,000) using enhanced chemiluminescence substrate (PerkinElmer, NEL105001EA) on a Li-COR Odyssey system for digital acquisition. Gel source data are shown in Supplementary Fig. 1.

Carboxypeptidase protection assay of Mfn2. Mitochondrial membranes were prepared⁶ from mouse hearts overexpressing human Mfn2¹³ and digested with 200 ng ml⁻¹ each of carboxypeptidases A and B (Sigma-Aldrich) for increasing times at room temperature, followed by immunoblotting using an antibody (Abcam ab56889) specific to human Mfn2 HR2.

Live-cell imaging. MEFs or neurons were grown on chamber slides and infected with recombinant adenovirus for 48–72 h unless otherwise indicated. MEFs were stained with MitoTracker Green (200 nM, Invitrogen M7514) and Hoechst (10 mg ml⁻¹, Invitrogen H3570), with or without tetramethylrhodamine ethyl ester (TMRE; 200 nM, Invitrogen T-669), which stains normally polarized mitochondria. Neuronal mitochondria were labelled with adenoviral-expressed mitoGFP plus TMRE, or with adeno-mitoDsRed for time-lapse studies. Cover slips were loaded onto a chamber (Warner instrument, RC-40LP) in modified Krebs–Henseleit buffer (138 mM NaCl, 3.7 mM KCl, 1.2 mM KH₂PO₄, 15 mM Glucose, 20 mM HEPES and 1 mM CaCl₂) at room temperature and visualized on a Nikon Ti Confocal microscope equipped with a 60× 1.3 NA oil-immersion objective. For mitophagy studies, MEFs were stained with MitoTracker Green and transfected with mCherry–Parkin on an adenovirus vector or stained with LysoTracker Deep Red (50 nM, Invitrogen L-7528); carbonyl cyanide-*p*-trifluoromethoxyphenylhydrazone (FCCP, 10 µM for 1 h) was applied as a positive control. Mitochondrial fusion was measured 2 and 6 h after the PEG-mediated cell fusion of MEFs treated 48 h previously with adenovirus-transfected mitoGFP or adeno-mitoDsRed¹⁴. Laser confocal fluorescence was excited with 561 nm (for MitoTracker Green, GFP, FITC), 637 nm (for TMRE, LysoTracker Red, mCherry–Parkin, MitoTracker Orange) or 408 nm (for Hoechst) laser diodes.

Image analysis. Mitochondrial aspect ratio was calculated using automated edge detection and Image J software as described⁶. Mitochondrial depolarization was calculated as the percentage of green mitochondria visualized on MitoTracker Green and TMRE merged images, expressed as green/(green + yellow mitochondria) × 100. Parkin aggregation was calculated as the percentage of cells with mitochondrial clumping of mCherry–Parkin⁶. Lysosomal engulfment of mitochondria was identified by the co-localization of LysoTracker Red and MitoTracker Green⁶.

TAT-minipeptide design, synthesis and use. We designed three polypeptides derived from MFN2: two from the HR1 region flanking its REQQ bend and one from the adjacent C-terminal area between HR1 and the transmembrane domain. Since these peptides are derived from α-helical domains, we attempted to conserve their secondary structure by maintaining a length of ~20 amino acids, which was predicted to create a stable α-helix of about five turns. We considered that the native ~20 amino acid polypeptides might not effectively compete with the endogenous HR1–HR2 interactions, as each represents only a portion of the overall HR1-interacting domain. We therefore generated two analogues for each of the three polypeptides, substituting the mid-peptide Leu residues for either Gly or Pro. We predicted that substituting Gly for the central Leu residues would increase the rotational flexibility of the peptide and enhance its ability to compete for, and interfere with, endogenous inter-domain interactions in intact Mfn2. Conversely, we reasoned that peptides in which Pro was substituted for the central Leu residue would be less active as the α-helix would be kinked, thus limiting competition for endogenous interactions. The nine polypeptides were expressed using adenovirus as mini-genes for screening and initial characterization. The two biologically active polypeptides, MFN2 MP1^{Gly} and MP2^{Gly}, were chemically synthesized and introduced into cells using TAT_{47–57} conjugation, which enabled transmembrane delivery²¹. Except for dose-response studies, 1 mM stock concentrations of TAT-minipeptides in sterile water were applied to cultured MEFs at a dilution factor of 1:1,000 or 1:200 to achieve final concentrations of 1 or 5 µM.

Parent peptide sequences (the central Leu that was mutationally substituted is underlined) are:

MFN2 367–384 (MP1): DIAEAVRLIMDSLHMAAR
MFN2 398–418 (MP2): QDRLKFIDKQLELLAQDYKLR
MFN2 428–448 (MP3): RQVSTAMAEIIRLSVLVDDY

GTPase assay. The effects of each minipeptide on Mfn2 GTPase activity were evaluated by incubating recombinant Mfn2 (100 ng; OriGene) with 1 µM control TAT peptide, TAT-MP1^{Gly}, or TAT-MP2^{Gly} for 15 min before the addition of 0.5 mM GTP for 1 h at 37°C. GTPase activity, relative to no-peptide control, was measured using a colorimetric GTPase assay kit (Novus Biologicals) following the manufacturer's instructions.

Minipeptide-Mfn2 binding assay. Mfn2 was bound to carboxylic acid sites of AGILE (Nanomaterial Diagnostics) graphene chips²² as previously described²³. Minipeptides were applied at varying concentrations (500 nM–1 mM) and the change in sensor chip-charge was recorded for 15 min. Responses of 25 sensors on a single assay-chip were averaged. A Hill-equation fit was used to derive the *K_d*. All samples were identical before allocation of treatments and analyses were performed by an observer blinded to the experimental conditions.

Mfn2 FRET assay. mCerulean1 and mVenus reporter genes were cloned onto the 5' and 3' ends of human Mfn2 cDNA in pcDNA3.1 and confirmed by DNA sequencing. Lipofectamine-transfected biosensor-expressing HEK 293T cells were imaged on an Olympus IX81 ZDC inverted microscope under 40× magnification. mCerulean was excited at 436 nm with emission at 480 nm. mVenus was excited at 500 nm with emission at 535 nm. FRET was imaged with excitation at 436 nm and emission at 535 nm. Studies were performed after adding GTPγS (1 mM, Sigma-Aldrich) to suppress Mfn2 GTPase activity. Live-cell FRET data were acquired every 60 s for 15 min before and 90 min after addition of TAT-conjugated minipeptides (5 µM).

Computational modelling of Mfn2 structure. The I-TASSER (Iterative Threading ASSEmbly Refinement) hierarchical approach to protein structure was used to derive a structural model of MFN2 (ref. 24). Structural templates are identified from the PDB database (<http://www.rcsb.org/pdb/home/home.do>) through sequence-similarity searches (threading). The top ten solutions were based on the bacterial dynamin-like protein structures in the PDB (ID: 2J69). The confidence of each model is quantitatively measured by C-score, calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [–5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice versa. The C-score in the top solution was –1.14. Another measure of the quality of a predicted structure is its estimated template-modelling score. The template-modelling score is a scale for measuring structural similarity and is estimated based on the C-score and protein length following the correlation observed between these

qualities. A template-modelling score >0.5 indicates a model of correct topology and a template-modelling score <0.17 means a random similarity. The estimated template-modelling score of the top solution is 0.57 ± 0.15 . Energy minimization and analysis of the structure was performed with MAESTRO tools (Maestro, v10.5, Schrödinger, LLC). PYMOL (The PyMOL Molecular Graphics System. v1.7; Schrödinger, LLC) was used for preparing images and videos.

Statistical analyses. The experiments were not randomized. Sample-size for mitochondrial studies was established *a priori* based on published experiments having similar endpoints⁶; experimental variance was similar between study groups and cell lines. Data are reported as mean \pm s.e.m. Statistical comparisons (two-sided) used one-way ANOVA and Tukey's test for multiple groups or Student's *t*-test for paired comparisons. $P < 0.05$ was considered significant.

Data availability statement. All data generated or analysed during this study are included in this published article and its Supplementary Information files.

19. Sobieski, C., Jiang, X., Crawford, D. C. & Mennerick, S. Loss of local astrocyte support disrupts action potential propagation and glutamate release synchrony from unmyelinated hippocampal axon terminals *in vitro*. *J. Neurosci.* **35**, 11105–11117 (2015).
20. Wang, W. *et al.* The ALS disease-associated mutant TDP-43 impairs mitochondrial dynamics and function in motor neurons. *Hum. Mol. Genet.* **22**, 4706–4719 (2013).
21. Mochly-Rosen, D., Das, K. & Grimes, K. V. Protein kinase C, an elusive therapeutic target? *Nat. Rev. Drug Discov.* **11**, 937–957 (2012).
22. Mackin, C. & Palacios, T. Large-scale sensor systems based on graphene electrolyte-gated field-effect transistors. *Analyst* **141**, 2704–2711 (2016).
23. Qvit, N., Disatnik, M. H., Sho, E. & Mochly-Rosen, D. Selective phosphorylation inhibitor of delta protein kinase C–pyruvate dehydrogenase kinase protein–protein interactions: application for myocardial injury *in vivo*. *J. Am. Chem. Soc.* **138**, 7626–7635 (2016).
24. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).

a

hMfn1	1	MAE-PVSPLKHFVLAKKAITAIFDQLLFEVTEGSHFVEATYKNPELDRIATEDDDLDEMVG	59
hMfn2	21	MAE SPLKHFV AKK I IF+QL ++ E + F+E TY+N ELD + TE+ ++++G	80
hMfn1	60	YKDKLSIIGEVLSRRHMKVAFFGRTSSGKSSVINAMLWDKVLPSGIGHITNCFLSVEGTD	119
hMfn2	81	YLSKVRGISEVLARRHMKVAFFGRTSNGKSTVINAMLWDKVLPSGIGHTTNCFLRVEGTD	140
hMfn1	120	GDKAYLMTEGSDEKKS VKTVNQLAHALHMDKDLKAGCLVRVFWPKAKCALLRDDLVLDVS	179
hMfn2	141	G +A+L+TEGS+EK+S KTVNQLAHALH DK L AG LV V WP +KC LL+DDLVL+DS	200
hMfn1	180	PGTDVTELD SWIDKFCLDADVFVLVANSESTLMNTEKHFFHKVNERLSKPNI FILNNRW	239
hMfn2	201	PGIDVTELD SWIDKFCLDADVFVLVANSESTLMQTEKHFFHKVNERLSRPNI FILNNRW	260
hMfn1	240	DASASEPEYMEDVRRQHMERCLHFLVEELKVVNALEAQNRIFFVSAKEVLSARKQKAQGM	299
hMfn2	261	DASASEPEYME+VRRQHMERCLV+EL VV+ +A +RIFFVSAKEVL+AR QKAQGM	320
hMfn1	300	PESGVALAEGFHARLQEFQNFQIFEECISQSAVKTKEQHTIRAKQILATVKNIMDSVN	359
hMfn2	321	PE G ALAEGF R+ EFQNFQIFEECISQSAVKTKEQHT+RAKQI V+ IMDS++	380
hMfn1	360	LAAEDKRHSVEEREDQIDRLDFIRNQMNLLTLDVKKKIKETEEVANKVSCAMTDEICR	419
hMfn2	381	+AA +++ Y E RE++ DRL FI Q+ LL D K +IK++TEEV +VS AM +EI R	440
hMfn1	420	LSVLVDFCSEFHPNPDLVKIYKSELNKHIEDGMGRNLADRCTDEVNALVLQTQQEIIEN	479
hMfn2	441	LSVLVD++ +FHP+P VLK+YK+EL++HIE+G+GRN++DRC+ + + QQ++I+	500
hMfn1	480	LKPLLPAGIQDKLHTLIPCKKFDLSYNLNYHKLCSDFQEDIVFRFSLGWSSLVHRFLGPR	539
hMfn2	501	LKPLLP ++ ++ L+P + F L+Y+LN KLC+DFQEDI F FSLGW+ LV+RFLGP+	560
hMfn1	540	NAQVRLLGLSEPIFQLPRSLASTPTAPTTPATPDNA-SQEEMLITLVTLGLASVTSRTSMG	598
hMfn2	561	N++R L+G ++ Q+ R + TP P+ P P + +QEE M+++VTGLAS+TSRTSMG	617
hMfn1	599	NSRRALMGYND---QVQRPIPLTPANPSMPLPQGS LTQEEFMVSMVTGLASLTSTRTSMG	658
hMfn2	618	IIIVGGVIWKTIGWKL SVSLTMYGALYLYERLSWTHAKERAFKQQFVNYATEKLRMIV	677
hMfn1	659	I++VGGV+WK +GW+L+++S +YG LY+YERL+WTT AKERAFK+QFV +A+EKL+++	718
hMfn2	678	ILVVGCVVWKAVGWRLIALSFGLYGLLYVYERLTWTTAKERAFKRQFVEHASEKLQLVI	737
hMfn1	719	SSTFANCSSHQVKKQIATTFARLCQQVDITQKQEEFIARLPKEIDQLEKIQNNSKLLRNK	
hMfn2	738	S T +NCSHQV+Q+++ TFA ICQQVD+T++ LE+EIA + K+I+ L+ +Q+ +KLLRNK	
hMfn1	719	SYTFSNCSHQVQQETSGTFAHLCQQVDVTRENTQEIAANKKTEVLDLSQSKAKLLRNK	
hMfn1	719	AVQLENELENFTKQILPSS	737
hMfn2	738	A L++EL FT Q+L S	756
hMfn1	719	AGWLDSEINMFTHQYLQPS	

b

Human Mfn2 inter-molecular HR2-HR2 interacting residues

hMfn2 -HR2 NH₃⁺ -GSNCSSHQVQQELSGTFAHLCQQVDVTRENTQEIAANKKTEVLDLSQSKAKL-LRNKAGW-LDSELNMFTHQYLQPS-COO⁻

hMfn2 -HR2 COO⁻ -SPLQHTMNEESD-LW-GAKNRLK-LKSKQSDVE-LKKNMAA-EQENERLVDVQQCHAFGTGSLEQQVQHSCNSG-NH₃⁺

c

Human Mfn2 intra-molecular HR1-HR2 interacting residues

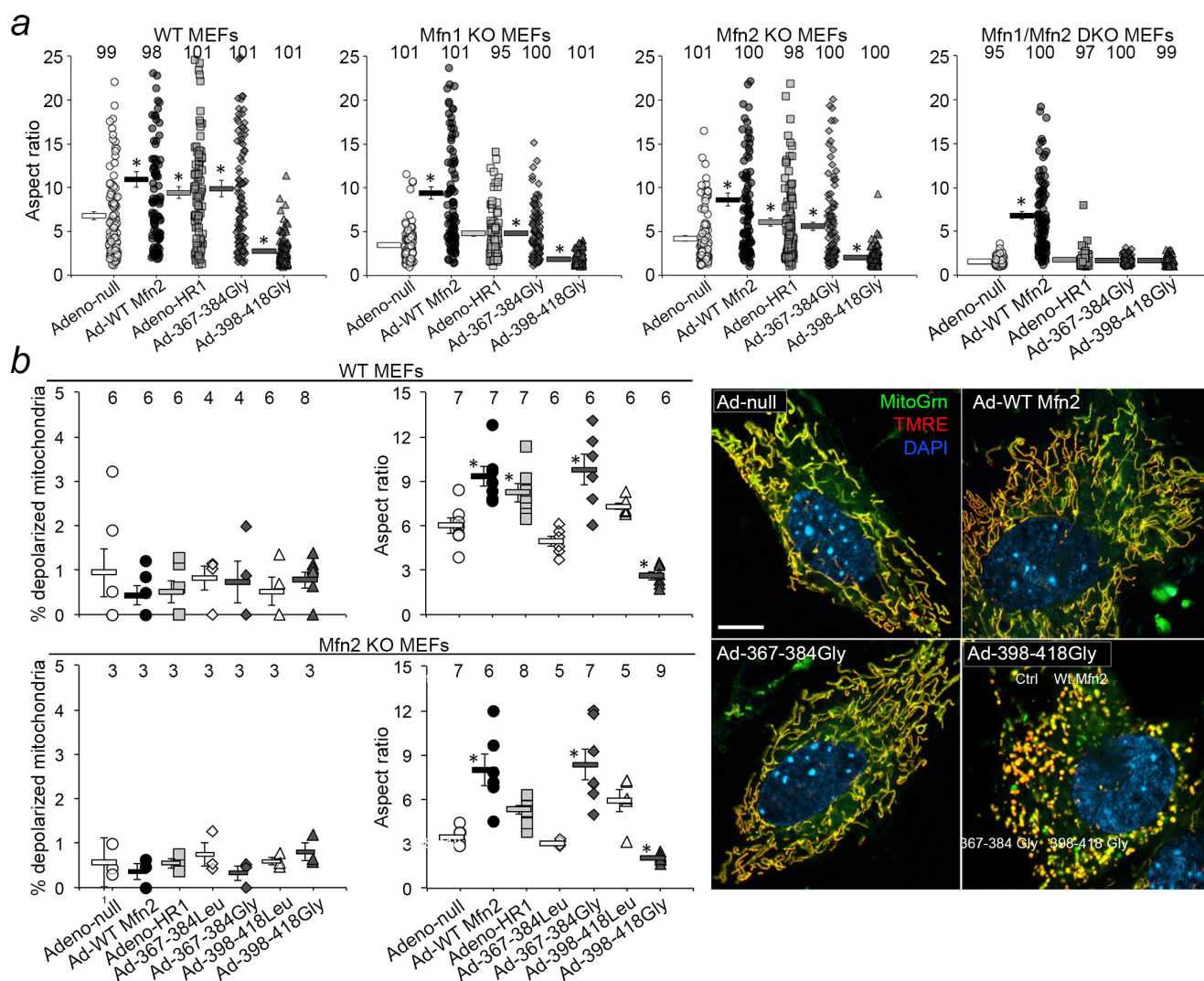
hMfn2 - HR1 NH₃⁺ -FQNFERRFEECISQSA--V--KTKFEQHTVRAKQI-EEA-VRLIMDSHMAAREQQVY--EEMREERDRKFDKQLELLAQDKLRKQ-COO⁻

hMfn2 - HR2 COO⁻ -SPQLQHTFMNEESD--LW--GAKNRLKAKSKQSDVE-LKKNMAAIEQENERLVDVQQCHAFGTGSLEQQVQHSCNSG-NH₃⁺

Extended Data Figure 1 | Interacting domains in human mitofusins.

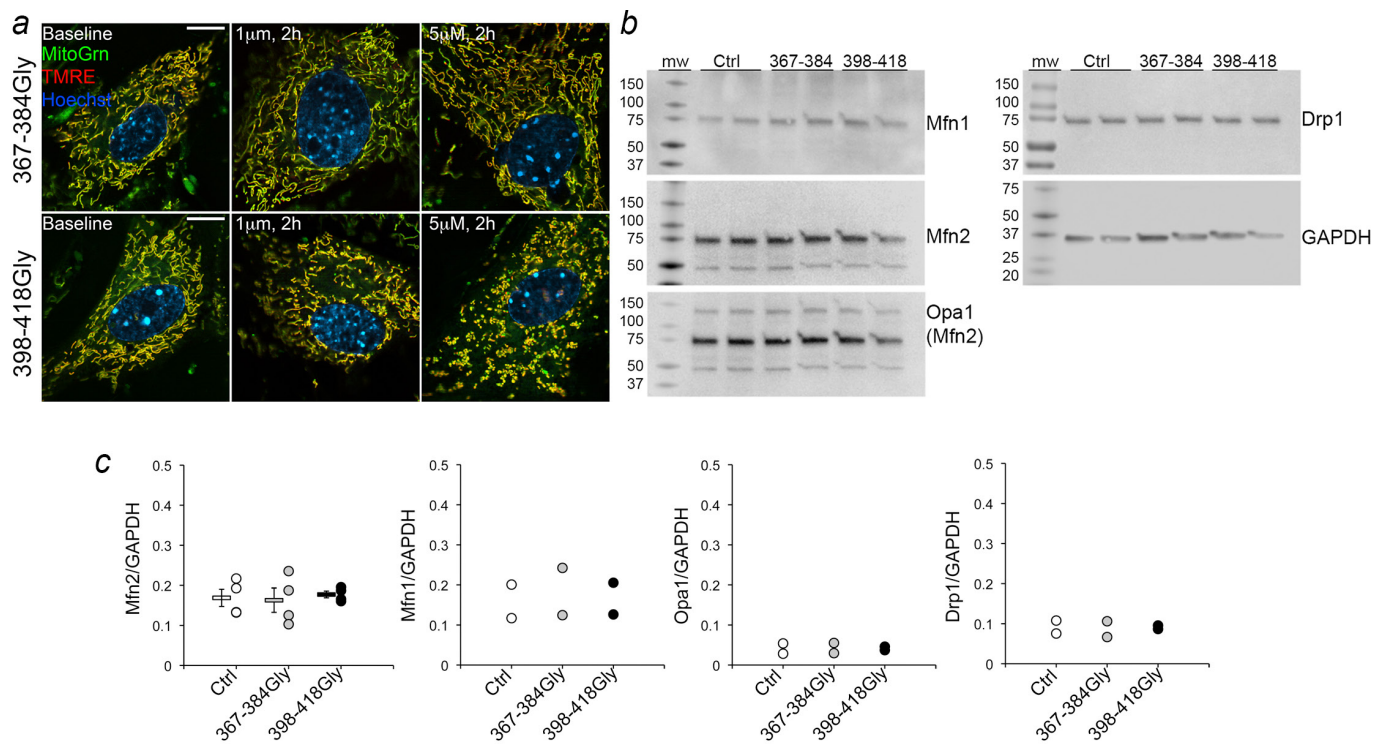
a. Amino-acid alignment of human MFN1 and MFN2. Conservative amino-acid substitutions are indicated by plus symbols. MFN domains are colour-coded as in Fig. 1a: purple is the GTPase domain, red is HR1, black is transmembrane, green is HR2. Yellow amino acids in HR1 and

red amino acids in HR2 are residues central to inter- or intramolecular binding. Areas where local unfolding of HR1 and HR2 α -helices occurs are highlighted yellow-green. b. Antiparallel intermolecular MFN2 HR2-HR2 binding as predicted in ref. 8. c. Intramolecular MFN2 HR1-HR2 binding as depicted in Fig. 1c and Supplementary Video 2.



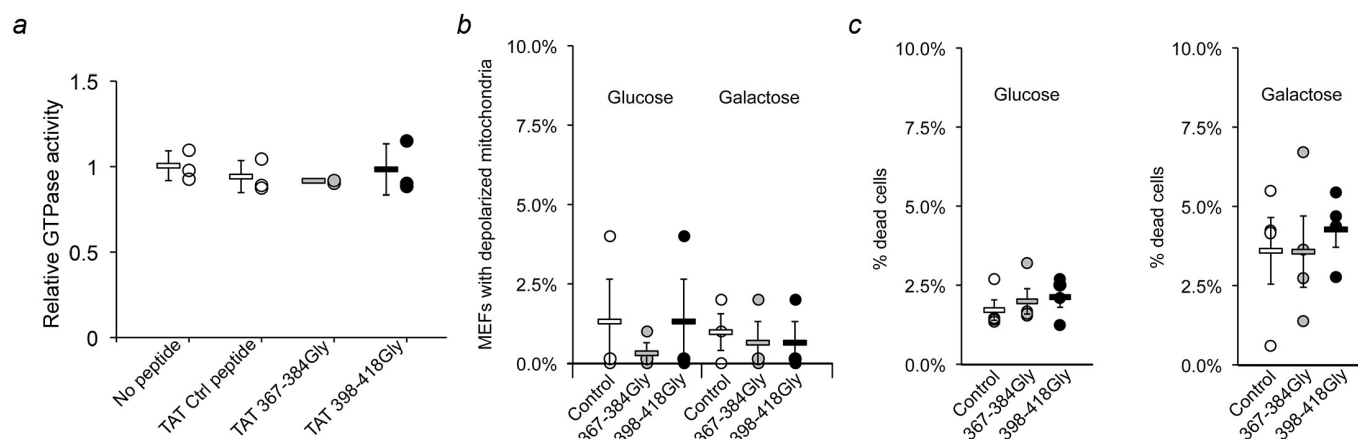
Extended Data Figure 2 | Adenovirally-expressed Gly-substituted mini-peptides derived from MFN2 HR1 specifically regulate Mfn-mediated mitochondrial fusion. **a**, Changes in mitochondrial aspect ratio provoked by the different Mfn expression profiles of MP1^{Gly} (adeno-367-384Gly) and MP2^{Gly} (adeno-398-418Gly) in MEFs, compared to MEFs transfected with wild-type *Mfn2* (ad-WT *Mfn2*) and MP-HR1 (adeno-HR1). Each point is the aspect ratio of an individual mitochondrion; group data are mean \pm s.e.m.; * $P < 0.05$ versus MEFs transfected with empty adenovirus (adeno-null) by ANOVA. Exact numbers of mitochondria measured from four or five separate experiments per condition are shown. **b**, Mitochondrial polarization status and aspect ratio in wild-type and *Mfn2*-knockout MEFs expressing

wild-type MFN2 or fragments thereof. Each point is the mean of approximately 20 mitochondria from approximately 5 cells in the number of independent experiments indicated on the graph. Group data are mean \pm s.e.m.; * $P < 0.05$ versus MEF mitochondria transfected with empty adenovirus vector by ANOVA. Representative (of 15–40 images per group) merged confocal images of wild-type MEFs transfected with *MFN2*, or the two biologically active peptides are to the right. Scale bar, 10 μ m. **c**, MitoTracker Green and mCherry Parkin (top) or Lyso-Red (bottom) co-stained wild-type or *Mfn2*-knockout MEFs before (left) or 1 h after (right) treatment with 50 nM FCCP. Group data are mean \pm s.e.m.; there were no significant differences between groups (ANOVA; $n = 3$ independent experiments per condition). Scale bars, 10 μ m.



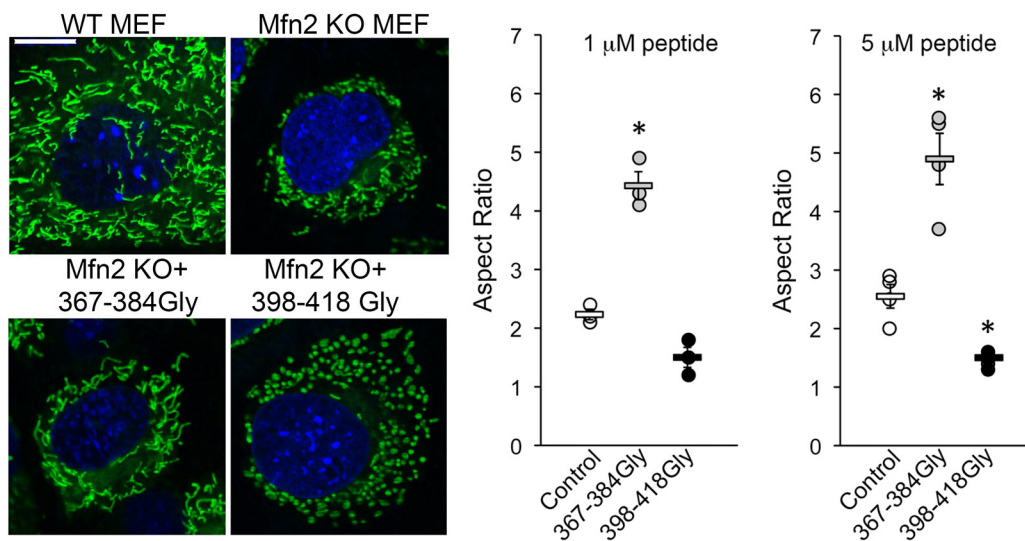
Extended Data Figure 3 | Effects of TAT-conjugated minipeptides on MEFs. a, Merged confocal images (representative of 30) of mitochondria in MEFs 2 h after application of TAT-MP1^{Gly} (367-384Gly; top) or TAT-MP2^{Gly} (398-418Gly; bottom). Complete dose-response relationship is in Fig. 2d. **b,** Immunoblot analysis of mitochondrial dynamics proteins 4 h after application of TAT-MP1^{Gly} (367-384) or TAT-MP2^{Gly} (398-418).

Opa1 is the inner mitochondrial membrane fusion protein Optic atrophy 1; Drp1 is the fission protein Dynamin-related protein 1. Opa1 immunoblot was co-probed for Mfn2. **c,** Immunoblot quantification. Data are mean \pm s.e.m.; $n = 4$ for Mfn2; $n = 2$ for all other proteins. Uncropped original blots are in supplementary information Fig. 1.



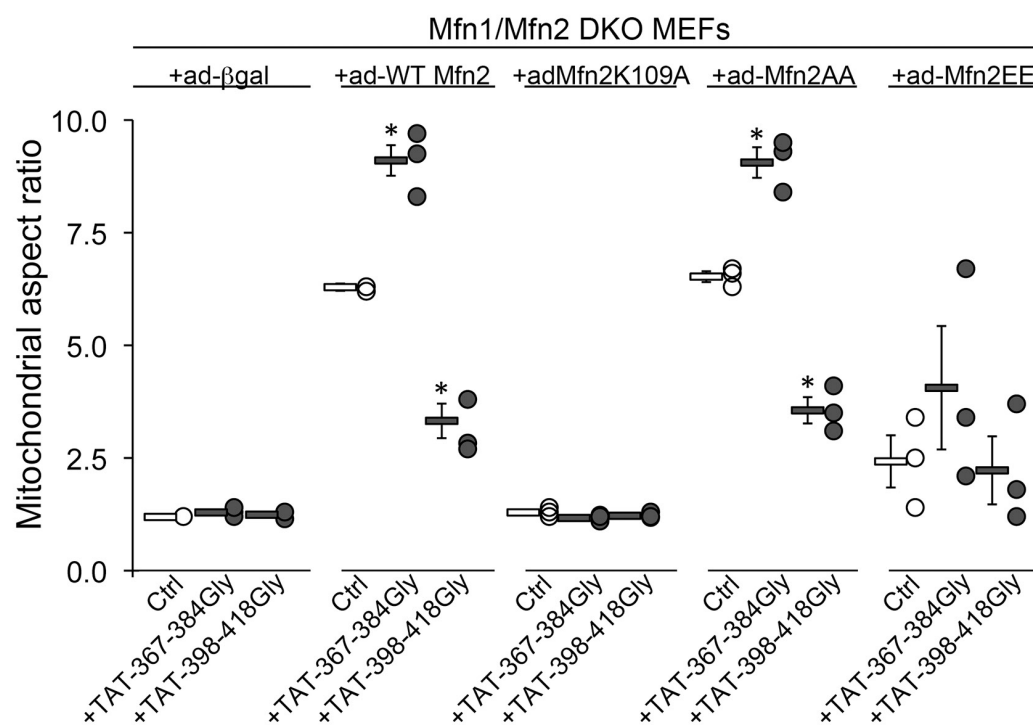
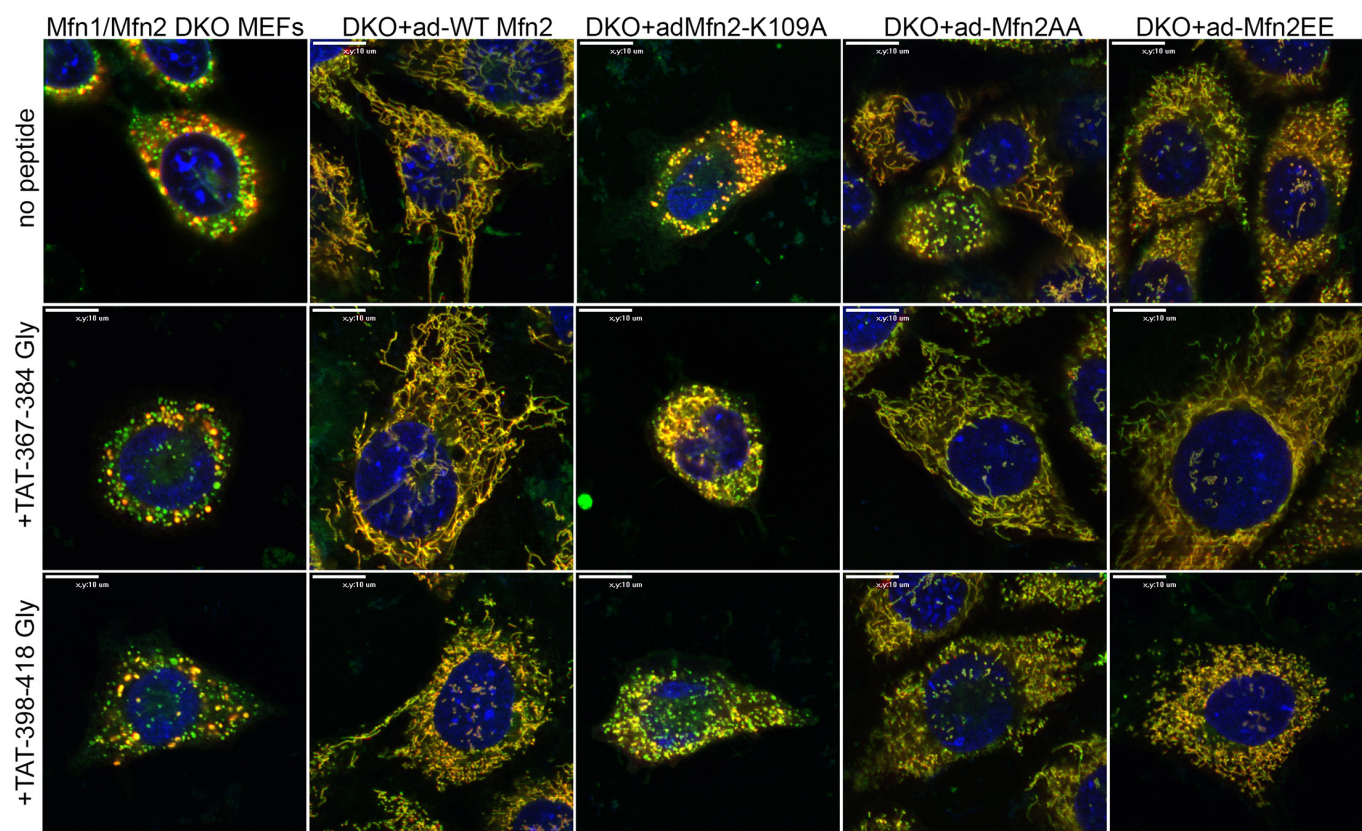
Extended Data Figure 4 | MFN2 HR1-derived TAT-mini-peptides do not suppress GTPase activity, impair mitochondrial polarization, or increase cell death. **a.** Relative GTPase activity of recombinant MFN2 is shown with respect to a no-Peptide control. Background signal from no GTP blank was subtracted from each value before normalization. Data are mean \pm s.d. of three independent experiments. **b, c.** Mitochondrial depolarization was assessed as loss of TMRE staining (**b**) and cell death

was assessed as ethidium bromide uptake (**c**) in wild-type MEFs cultured in 4,500 mg l⁻¹ glucose (left) or galactose (right) treated with or without 1 μ M TAT-mini-peptide for 24 h. $n = 3$ or 4 independent experiments per condition. All group data in panels **a–c** are mean \pm s.e.m. There were no significant differences between TAT-mini-peptide treatment groups (ANOVA).



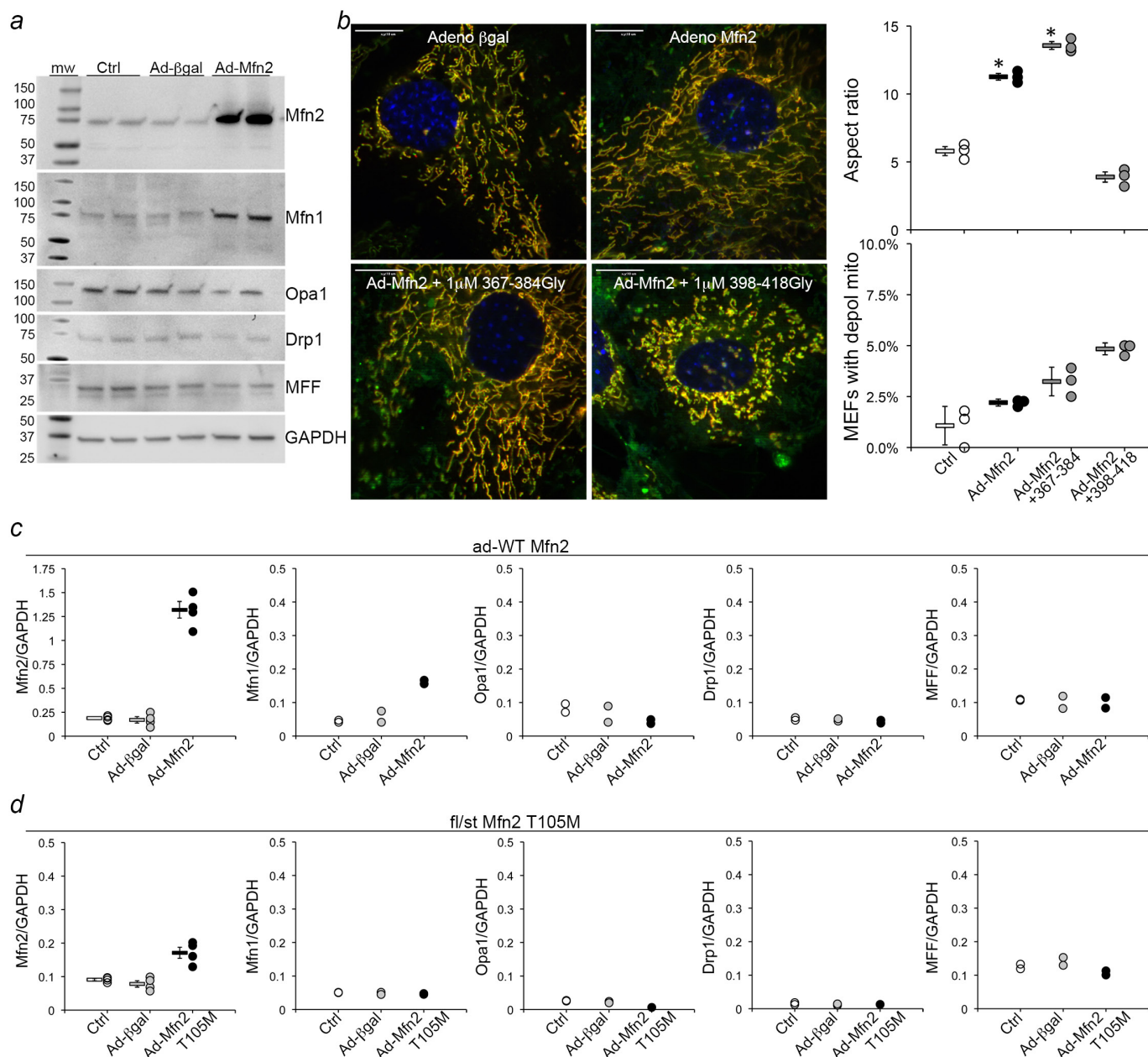
Extended Data Figure 5 | TAT-MP1Gly rescues mitochondrial dysmorphology in a cell model of autosomal recessive CMT2A. Representative (of 15–20 per group) confocal micrographs of *Mfn2*-knockout MEFs with stained mitochondria (green) and nuclei (blue)

knockout (KO) at baseline or 24 h after application of mini-peptide (left); a representative normal MEF is shown for comparison. Mean normal MEF mitochondrial aspect ratio is 6.2. Group data are mean \pm s.e.m.; * $P < 0.05$ versus control (ANOVA), $n = 3$ or 4 per group. Scale bar, 10 μ m.



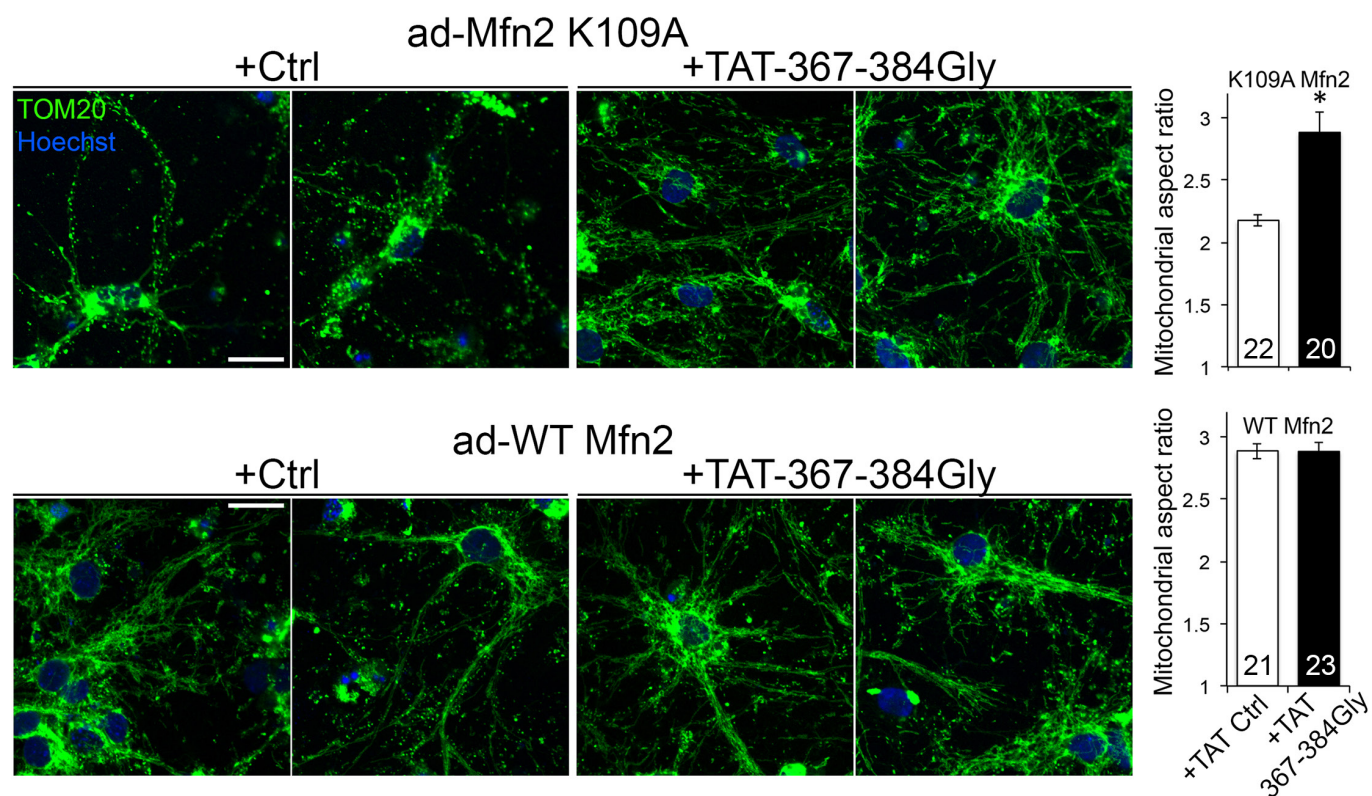
Extended Data Figure 6 | Minipeptide effects on dysfunctional MFN2 mutants. Wild-type MFN2 or the indicated mutants were expressed for 48 h using adenovirus vectors (ad) in Mfn-null MEFs (*Mfn1/Mfn2* double knockout (DKO)). TAT-conjugated minipeptides or vehicle (sterile water) were added and mitochondrial morphometry assessed 24 h later. Images

(representative of 15 per group) are of merged MitoTracker Green (green) and TMRE (red; stains active, polarized mitochondria) channels. Scale bars, 10 μ m. Quantitative data for aspect ratio are shown below ($n = 3$ per condition). Group data are mean \pm s.e.m.; * $P < 0.05$ versus same-group Ctrl (ANOVA).



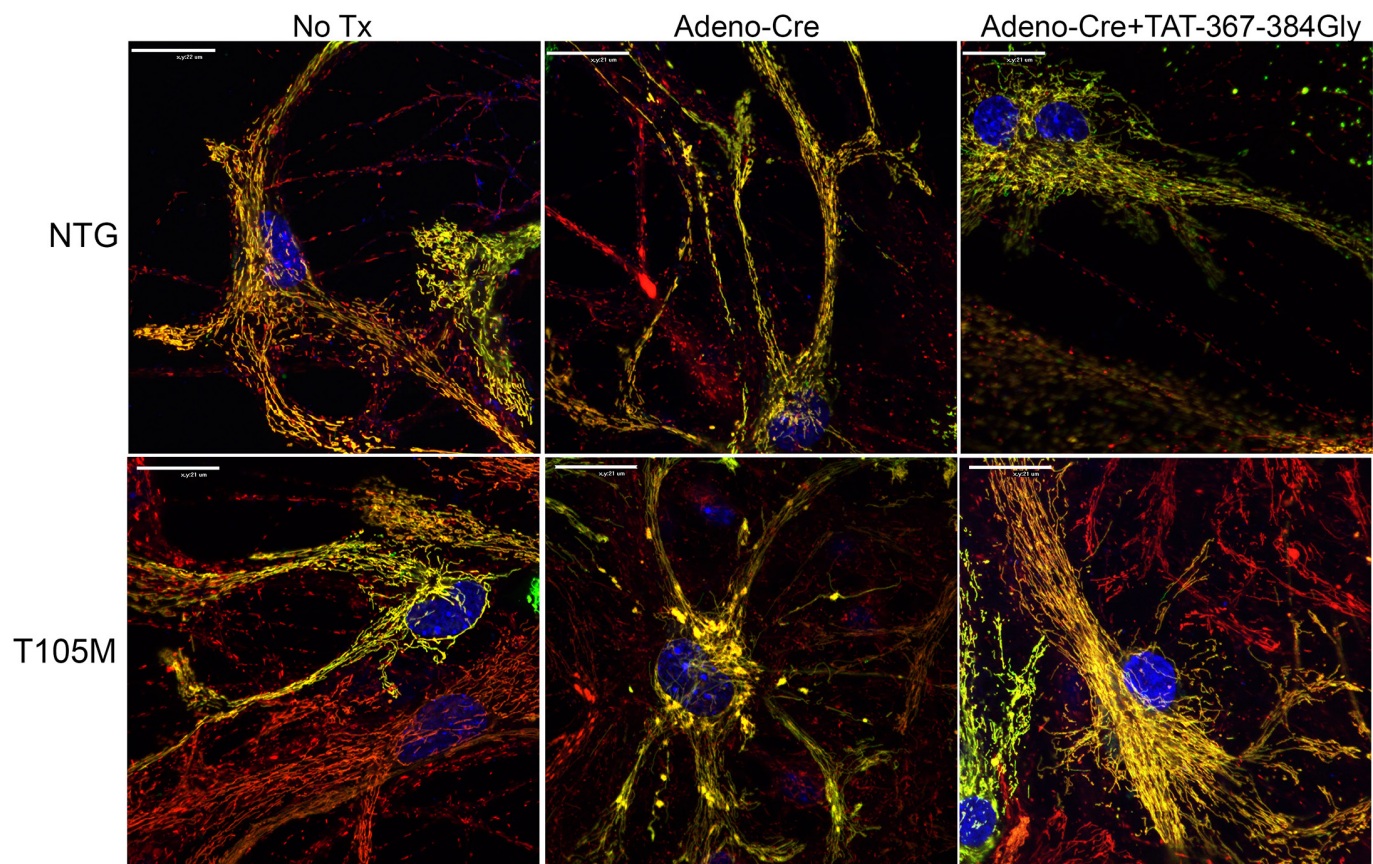
Extended Data Figure 7 | Wild-type MFN2 expression to parallel MFN2^{T105M} expression studies. **a**, Immunoblot analysis of mitochondrial dynamics proteins in cultured MEFs transduced with adenovirus-transfected β gal (viral control) or adenovirus-transfected human MFN2. MFN2 expression increased fivefold over uninfected Ctrl; apparent increased Mfn1 signal is cross immunoreactivity with MFN2 as shown by slightly faster migration. **b**, Merged confocal images (MitoTracker Green and TMRE; representative of 15 per group) of MEFs expressing

wild-type MFN2 and treated with TAT-mini-peptides as shown. Quantitative data ($n = 3$ independent experiments) are on the right. Group data are mean \pm s.e.m.; $*P < 0.05$ versus control by one-way ANOVA. Scale bars, 10 μ m. **c**, Immunoblot quantifications for panel **a**. **d**, Immunoblot quantifications for Fig. 4b, for comparison. $n = 4$ for MFN2; $n = 2$ for all other proteins. y axis scales are identical except MFN2 in **c**, which is expanded to accommodate high ($5 \times$ normal) ad-WT MFN2 expression level.



Extended Data Figure 8 | TAT-MP1^{Gly} corrects mitochondrial fragmentation provoked by Mfn2^{K109A} in cultured rat motor neurons. Representative (of >20 per group) anti-TOM20/Hoechst images of formalin-fixed cultured neurons transfected with genes encoding

Mfn2^{K109A} (top) or wild-type Mfn2 (bottom) and treated for 24 h with 1 μ M TAT control (Ctrl) or TAT-MP1^{Gly} minipeptide. Scale bars, 10 μ m. Experimental *n* number per treatment group is shown on the graph. Data are mean \pm s.e.m. **P* < 0.05 versus Ctrl (Student's *t*-test).



Extended Data Figure 9 | TAT-MP1^{Gly} reverses mitochondrial pathology in cultured *MFN2*^{T105M} mouse neurons. Live-cell confocal imaging of MitoGFP (green), TMRE (red) and Hoechst (blue nuclei) stained mouse hippocampal neurons (representative of >30 images per group). Normal control neurons from nontransgenic (NTG) mouse pups

are on top; *MFN2*^{T105M} *fl/st* transgenic mouse neurons are on bottom. Adeno-MitoGFP was added at low titers to enhance visualization of individual neurons. Note mitochondrial fragmentation and clumping after Cre-induced induction of *MFN2*^{T105M}, and reversal after 24 h of TAT-MP1^{Gly} treatment. Scale bars, 20 μm.

The pathway to GTPase activation of elongation factor SelB on the ribosome

Niels Fischer¹, Piotr Neumann², Lars V. Bock³, Cristina Maracci⁴, Zhe Wang⁵, Alena Paleskava^{4†}, Andrey L. Konevega^{4†}, Gunnar F Schröder^{5,6}, Helmut Grubmüller³, Ralf Ficner², Marina V. Rodnina⁴ & Holger Stark¹

In all domains of life, selenocysteine (Sec) is delivered to the ribosome by selenocysteine-specific tRNA (tRNA^{Sec}) with the help of a specialized translation factor, SelB in bacteria. Sec-tRNA^{Sec} recodes a UGA stop codon next to a downstream mRNA stem-loop. Here we present the structures of six intermediates on the pathway of UGA recoding in *Escherichia coli* by single-particle cryo-electron microscopy. The structures explain the specificity of Sec-tRNA^{Sec} binding by SelB and show large-scale rearrangements of Sec-tRNA^{Sec}. Upon initial binding of SelB-Sec-tRNA^{Sec} to the ribosome and codon reading, the 30S subunit adopts an open conformation with Sec-tRNA^{Sec} covering the sarcin-ricin loop (SRL) on the 50S subunit. Subsequent codon recognition results in a local closure of the decoding site, which moves Sec-tRNA^{Sec} away from the SRL and triggers a global closure of the 30S subunit shoulder domain. As a consequence, SelB docks on the SRL, activating the GTPase of SelB. These results reveal how codon recognition triggers GTPase activation in translational GTPases.

Protein synthesis on the ribosome is promoted by several GTPases. These include the universal elongation factor EF-Tu (in bacteria, or eEF1A in eukaryotes) and a specialized factor SelB¹ (eEFSec-SBP2 in eukaryotes²). EF-Tu and SelB deliver aminoacyl-tRNA (aa-tRNA) to the ribosome at the cost of GTP hydrolysis^{3,4}. In contrast to EF-Tu, which is a carrier for all canonical aa-tRNAs, SelB is specialized to bind only the tRNA that is specific for Sec^{5,6}. Sec-tRNA^{Sec} recodes a UGA stop codon upstream of a structured mRNA stem-loop, the Sec insertion sequence (SECIS)⁷. EF-Tu and SelB share evolutionary conserved domains 1–3, of which domain 1 binds GTP^{5,8–11}. SelB has an additional domain 4 (winged helix motifs 1–4) that recruits the factor to the SECIS¹⁰. Sec has unique chemical properties, which make it an essential residue in the active site of key redox enzymes in all domains of life². Defects in Sec biosynthesis and selenoprotein malfunction have been linked to an increasing number of diseases in humans¹². Sec incorporation into non-selenoproteins, as well as the exploitation of the Sec machinery to insert designer amino acids, has received growing interest for biotechnological applications^{13,14}. However, the structural basis for UGA recoding by SelB-Sec-tRNA^{Sec} has not been elucidated yet. Here we use single-particle cryo-electron microscopy (cryo-EM) and extensive sorting of cryo-EM images to obtain snapshots of intermediates along the pathway of Sec-tRNA^{Sec} delivery to the ribosome. The cryo-EM structures, in combination with ensemble rapid kinetics and all-atom explicit-solvent molecular dynamics simulations, provide detailed insights into the mechanism of UGA recoding by Sec and suggest a universal mechanism for GTPase activation on the ribosome.

Visualizing intermediates of UGA recoding

We reconstituted recoding complexes *in vitro* with fMet-tRNA^{fMet} in the peptidyl site and SelB-Sec-tRNA^{Sec} stalled by the guanosine triphosphate analogue GDPNP at the aminoacyl site of the ribosome. As a model mRNA directing Sec incorporation, we used a construct with a AUG start codon and a UGA stop codon, followed by the SECIS of the

formate dehydrogenase H (*fdhF*) gene^{10,15}. We recorded ~1,000,000 cryo-EM particle images and applied extensive hierarchical computational classification to sort the particle images into structurally homogeneous sub-groups (Extended Data Fig. 1 and Methods). We obtained six distinct ribosome structures at 3.4–5.3 Å resolution, which depict the key intermediates of UGA recoding (Fig. 1a, Extended Data Fig. 2, Extended Data Table 1) and used the 3.4 Å map to build *de novo* atomic models for the SelB-GDPNP-Sec-tRNA^{Sec} ternary complex on the ribosome (Fig. 1b–e).

The initial complex depicts the ribosome complex before SelB-GDPNP-Sec-tRNA^{Sec} binding, with the SECIS flexibly bound at the mRNA entry channel, the UGA stop codon in the aminoacyl site and fMet-tRNA^{fMet} bound to the peptidyl site (Fig. 1a). We also obtained complexes in the transient states (reconstructed from only ~1% of images, Fig. 1f) of the codon-independent initial binding of SelB-GDPNP-Sec-tRNA^{Sec} to the SECIS and of codon reading with tRNA^{Sec} approaching the UGA codon. Although the overall structure of the two complexes is rather similar, with an overall root-mean-square deviation (r.m.s.d.) of 1.2 Å (Extended Data Fig. 1d), there are substantial local differences in functionally important regions, for example, an up to ~14 Å shift in the tRNA^{Sec} position. The highest resolution is provided by the 3.4 Å cryo-EM map of SelB-GDPNP-Sec-tRNA^{Sec} stalled after codon recognition in the GTPase-activated state (Fig. 1a–d). We also obtained reconstructions of two states after SelB dissociation and Sec-tRNA^{Sec} accommodation in the aminoacyl site, in which the tRNAs are present in either classical or hybrid states¹⁶.

Recognition of Sec-tRNA^{Sec}

SelB binds Sec-tRNA^{Sec} with extraordinary high affinity and discriminates against Ser-tRNA^{Ser}, which is an on-pathway intermediate of Sec-tRNA^{Sec} biosynthesis^{4,17}, and all other aa-tRNAs. Conversely, EF-Tu, which binds all elongator aa-tRNAs with uniform affinity, has very low affinity for Sec-tRNA^{Sec} (ref. 17). Comparison of the present SelB–

¹Department of Structural Dynamics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. ²Department of Molecular Structural Biology, Institute for Microbiology and Genetics, GZMB, Georg-August University Göttingen, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany. ³Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. ⁴Department of Physical Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. ⁵Institute of Complex Systems (ICS-6), Forschungszentrum Jülich, 52425 Jülich, Germany. ⁶Physics Department, Heinrich-Heine Universität Düsseldorf, 40225 Düsseldorf, Germany. [†]Present addresses: Molecular and Radiation Biophysics Department, B.P. Konstantinov Petersburg Nuclear Physics Institute of National Research Centre 'Kurchatov Institute', 188300 Gatchina, Russia (A.P., A.L.K.); St Petersburg Polytechnic University, Polytechnicheskaya, 29, 195251 St Petersburg, Russia (A.P., A.L.K.).

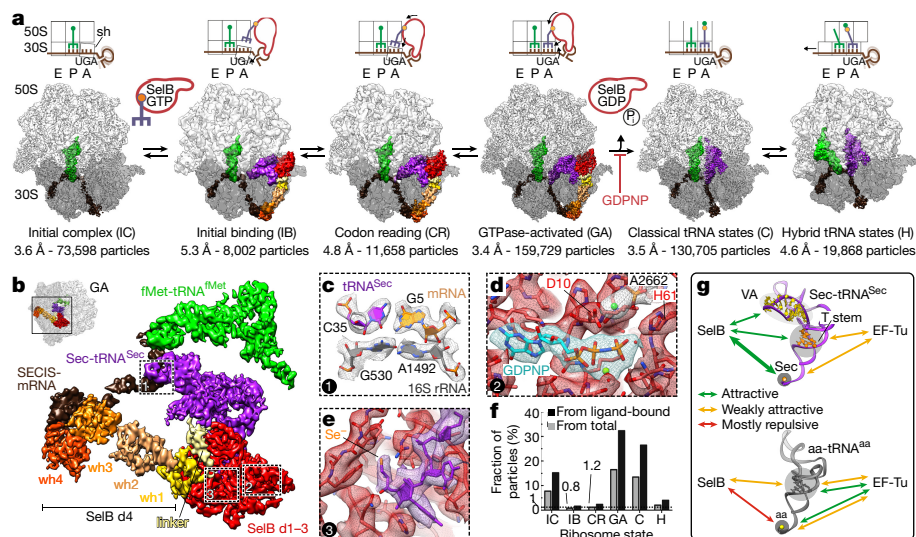


Figure 1 | Structures of intermediates along the pathway of UGA recoding by SelB-Sec-tRNA^{Sec}. **a**, Overview of the structures. Top row: Schematics of complexes with tRNA binding sites of the ribosome (A, P, E), UGA stop codon; sh, 30S subunit shoulder. Bottom row, cryo-EM reconstructions of distinct intermediates with the resolution (in Å) and the number of particles for each state. IC, initial complex; IB, initial binding; CR, codon reading; GA, GTPase-activated; C, classical tRNA; and H, hybrid tRNA states. **b**, Structure of SelB-GDPNP-Sec-tRNA^{Sec} and the

SECIS on the ribosome in the GTPase-activated state at 3.4 Å resolution. **d**, domains of SelB; wh, winged helix motifs. Numbers and dashed boxes indicate the regions depicted in **c**–**e**. **c**, Recognition of the UGA codon by the anticodon of tRNA^{Sec}. **d**, The GTPase centre of SelB in the activated state. **e**, Recognition of the selenol group (Se⁻) by SelB. **f**, Fraction of particles in each intermediate relative to the total number of particles (~1,000,000) or to the number of particles with ligand density (~500,000). **g**, Scheme of tRNA recognition by SelB versus EF-Tu. VA, variable arm.

GDPNP-Sec-tRNA^{Sec} structures with the known crystal structure of EF-Tu-GTP-Cys-tRNA^{Cys} (ref. 11) explains the specificity of tRNA^{Sec} recognition (Fig. 1g, Extended Data Fig. 3). The linker (Leu340–Gln363; *Escherichia coli* numbering used throughout) between domains 3 and 4 of SelB, which is absent in EF-Tu, binds and distorts the extra-long variable arm of tRNA^{Sec} (Extended Data Fig. 3a–c, l). Furthermore, an extended loop in SelB domain 3 (Ile326–Asn330) facilitates interactions with the acceptor- and T-stems of tRNA^{Sec} at the antiterminal box that distinguishes tRNA^{Sec} from canonical tRNAs¹⁸. The corresponding loop in EF-Tu adopts a different conformation that allows sequence-specific interactions with canonical tRNAs^{19,20}, but impedes interactions with tRNA^{Sec} (Extended Data Fig. 3e, f, j). Finally, the amino-acid-binding pocket of SelB is lined with positively charged residues (Extended Data Fig. 3g–i), allowing SelB to specifically recognize the negatively charged selenol group and to discriminate against Ser-tRNA^{Sec}. In EF-Tu the amino-acid-binding pocket is negatively charged; this, together with the low affinity of EF-Tu for the tRNA^{Sec} body²¹, explains why EF-Tu does not bind Sec-tRNA^{Sec}. Notably, the variable arm of tRNA^{Sec} undergoes a large-scale rotation of about 50° upon release of Sec-tRNA^{Sec} from SelB and accommodation in the aminoacyl site (Extended Data Fig. 3l). This rotation is essential to allow binding of the next aa-tRNA to the aminoacyl site after peptidyl-tRNA^{Sec} translocation to the peptidyl site.

SelB-specific domain 4

The SECIS recruits domain 4 of SelB¹⁰ and may guide the ternary complex towards the ribosome. We constructed a model for an early recruitment complex by docking SelB-Sec-tRNA^{Sec} onto the SECIS of the initial complex (Fig. 2a). In the model, SelB and Sec-tRNA^{Sec} do not interact with the ribosome, but a rotation would move SelB-Sec-tRNA^{Sec} into its position in the initial binding state. SECIS-dependent tethering via domain 4 is SelB-specific, as EF-Tu is recruited by the L7/12 stalk of the 50S ribosomal subunit²². In the subsequent states domain 4 serves as a flexible anchor for SelB-Sec-tRNA^{Sec} (Fig. 2b, Extended Data Fig. 4a, b). When SelB moves towards the GTPase-activated state, domain 4 maintains its interactions with the SECIS and the adjacent 30S elements (helix h16 of 16S ribosomal RNA (rRNA) and protein S4) and follows mainly the movements of the

30S subunit. These rigid-body motions of domain 4 do not affect the remaining part of SelB, because the highly flexible linker connecting domains 3 and 4 uncouples the two parts.

Requirements for SelB GTPase activation

GTP hydrolysis by EF-Tu is a major checkpoint that controls the rate and fidelity of decoding²³. Rapid GTP hydrolysis by EF-Tu is triggered by the correct codon–anticodon interaction²⁴; the requirements for the GTPase activation in SelB are unknown. In the presence of the correct SECIS and the UGA codon, the rate of GTP hydrolysis in SelB–[γ-³²P] GTP-Sec-tRNA^{Sec} on the ribosome is about 3 s⁻¹, which is four orders of magnitude higher than the intrinsic GTPase activity of the ternary complex in the absence of ribosomes (Fig. 2c, d, Extended Data Fig. 4c–f). Mutation of three key recognition bases in the SECIS^{10,25} results in a 600-fold reduction of the GTPase rate, in line with the role of the SECIS as a tether for SelB (see above and refs 10, 26). With a correct SECIS element, but a non-cognate UUC codon in the aminoacyl site, the rate of GTP hydrolysis is also reduced by 600-fold compared to the cognate complex (Fig. 2d), suggesting that recognition of the cognate UGA codon by Sec-tRNA^{Sec} is essential for GTPase activation. Thus, the requirement for the correct codon–anticodon interaction as a trigger for GTP hydrolysis is similar for SelB and EF-Tu. For EF-Tu, the existence of several pre-hydrolysis steps, including initial binding and codon reading, has been demonstrated by ensemble kinetics and single molecule FRET experiments^{27,28}. The similarities between SelB and EF-Tu as to the reaction pathway and codon-dependence of the GTPase activation suggest a conserved mechanism of decoding and aa-tRNA selection, with GTP hydrolysis being a crucial step between initial selection and proofreading^{24,29,30}.

Structural dynamics of the 30S subunit

Pioneering crystallographic studies on complexes of isolated 30S subunits demonstrated that cognate codon–anticodon complex formation triggers local rearrangements of three universally conserved bases of 16S rRNA in the decoding centre (G530, A1492, and A1493) and a global domain closure in the 30S subunit^{31,32}. Here we describe the rearrangements of the 30S subunit upon UGA codon recognition by Sec-tRNA^{Sec} in the context of the 70S ribosome (Fig. 3, Extended Data Fig. 5)

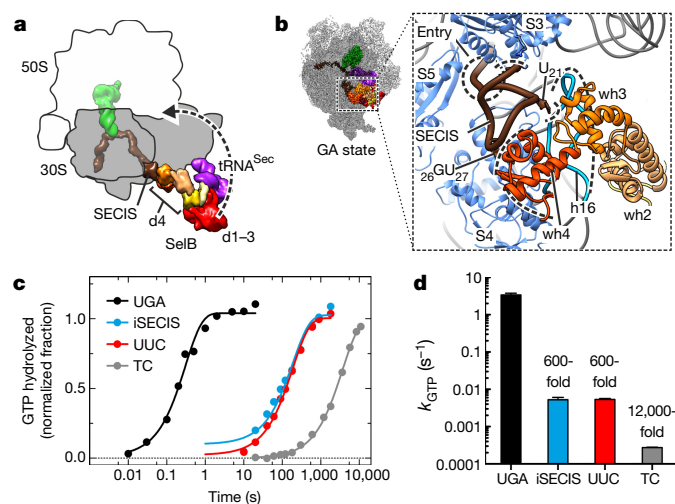


Figure 2 | Roles of SelB domain 4 and the GTPase activation on the ribosome. **a**, A model for the ribosome-independent recruitment of SelB onto the SECIS guiding SelB–Sec–tRNA^{Sec} to its position in the initial binding state (arrow). **b**, Interactions of SelB and the SECIS at the mRNA entry site (entry) in the GTPase-activated state marked by dashed contours; the interactions are maintained in the initial binding and codon reading states (data not shown). h16, helix of 16S rRNA; S3–S5, proteins S3 to S5. **c**, Time courses of GTP hydrolysis by SelB on ribosomes programmed with mRNA comprising a correct SECIS and a cognate UGA codon (UGA), or an incorrect SECIS and a cognate UGA codon (iSECIS), or a correct SECIS, but a non-cognate codon (UUC), and by the free SelB–GTP–Sec–tRNA^{Sec} (ternary complex; TC). **d**, The reduction in the GTP hydrolysis rates as compared to complexes carrying the cognate codon and correct SECIS (UGA); error bars, s.e.m. of the exponential fitting with $n = 2$ for UGA and $n = 3$ for iSECIS, UUC and ternary complex.

and show how these changes contribute to the mechanism of GTPase activation of SelB (Fig. 4).

In the initial complex before SelB–GDPNP–Sec–tRNA^{Sec} binding, the 30S subunit is in a non-rotated classical state^{33,34} (Fig. 3a, Extended Data Fig. 5d). The A1492 and A1493 bases appear to be in dynamic equilibrium between ‘flipped-in’ and ‘flipped-out’ states with respect to h44 of 16S rRNA. G530 is in the anti-conformation which is maintained in all subsequent states. Initial binding of SelB–GDPNP–Sec–tRNA^{Sec} is codon-independent, as evident from the large distance between tRNA^{Sec} anticodon and UGA codon (10 Å, Fig. 3a, Extended Data Fig. 5b). In the initial binding state, the flipped-in conformation of the A1492 and A1493 bases is stabilized and the aminoacyl site is open (Fig. 3a). The 30S shoulder rotates outwards, away from the 50S subunit. This 30S domain opening widens the intersubunit space, which may be important to facilitate initial binding of SelB domain 2 to the 30S shoulder or to prevent premature docking of SelB on the 50S subunit (see below). Further progression of tRNA^{Sec} towards the mRNA codon in the codon reading state may allow the interaction between the third codon base and the anticodon of tRNA^{Sec}, but this state is still incompatible with full codon–anticodon complex formation (Fig. 3a, Extended Data Fig. 5b). In the GTPase-activated state, the codon is recognized, the codon–anticodon helix adopts the Watson–Crick geometry, and A1492 and A1493 are stabilized in the flipped-out position, which results in a local closure of the aminoacyl site. Together with G530, the adenines interact with the minor groove of the codon–anticodon duplex at the first and second codon position. At the same time, the 30S subunit undergoes a large-scale domain closure, which mainly involves an inward rotation of the 30S shoulder and decreases the distance spanned by the SelB–GDPNP–Sec–tRNA^{Sec} complex. During all these steps, the universally conserved A1913 in helix 69 (H69) of 23S rRNA interacts with A1492 and A1493 in h44 in a state-specific fashion and acts to stabilize the flipped-in position of the two adenines in the initial binding and codon reading states (Extended Data Fig. 5a). The closed conformation of

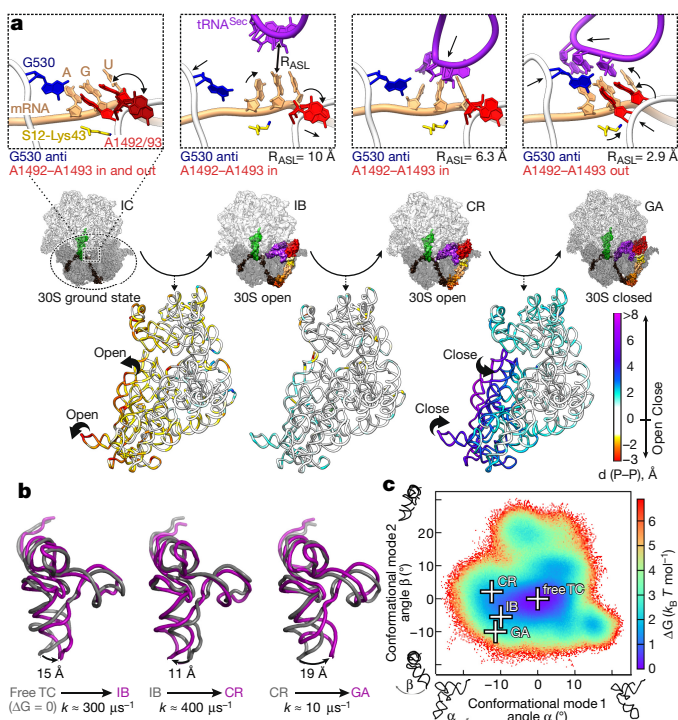


Figure 3 | Structural changes of 30S subunit and tRNA during UGA decoding. **a**, Coupling between local and global movements of the 30S subunit. Top row, local changes in the decoding centre. Middle row, overview of the states. Bottom row, global rearrangements of the 30S subunit. The heat map quantifies the displacement of 16S rRNA backbone phosphate atoms between consecutive states as obtained by superposition on 23S rRNA of the initial complex; negative values designate domain opening, positive values domain closure. R_{ASL} , distance between tRNA^{Sec} anticodon (N3 of C35) and UGA (N1 of G). **b**, Conformations of Sec–tRNA^{Sec}. Rearrangements from the free ternary complex SelB–GTP–Sec–tRNA^{Sec} to the initial binding state (left), from initial binding to codon reading (middle) and codon reading to GTPase-activated states (right). The deviations of the anticodon tip are measured in Å at the phosphate of C35. The free ternary complex structure and interconversion rates (k) were obtained by molecular dynamics simulation from **c**. **c**, Free-energy landscape of tRNA^{Sec} conformations in the free ternary complex obtained by 24-μs all-atom explicit-solvent molecular dynamics simulations (Methods). Crosses denote the tRNA conformations seen in the respective states on the ribosome; the free-energy minimum corresponds to the major conformer of the free ternary complex. The heat map denotes the free energy.

the 30S subunit in the GTPase-activated state is practically identical to that observed for the canonical EF–Tu–GDPNP–Trp–tRNA^{Trp} ternary complex on the ribosome³⁵. However, compared to the states preceding codon recognition, the rearrangement is more extensive in the present complexes than in those inferred from the structures of isolated 30S subunit complexes³¹ (Fig. 3a).

Conformational changes of tRNA^{Sec}

The aa-tRNA has an essential role in signalling rearrangements at the decoding site to the 50S subunit, leading to GTPase activation of EF–Tu^{36–38}. Our cryo-EM data show marked structural changes in tRNA^{Sec} on the ribosome, which affect three main areas of the tRNA^{Sec} molecule: the variable arm, the elbow region consisting of the D- and T-loops, and the D-stem–anticodon-stem region (Fig. 3b). The D- and T-loops move closer to each other in the initial binding state and change only slightly in other states. By contrast, the anticodon- and D-stems undergo large-scale distortions in each transition. In the GTPase-activated state tRNA^{Sec} adopts a conformation similar to that of the canonical aa-tRNAs in the EF–Tu–ribosome complex^{35,38}. In principle, the observed large-scale conformational rearrangements of the tRNA could present a kinetic barrier that controls the GTPase activation in

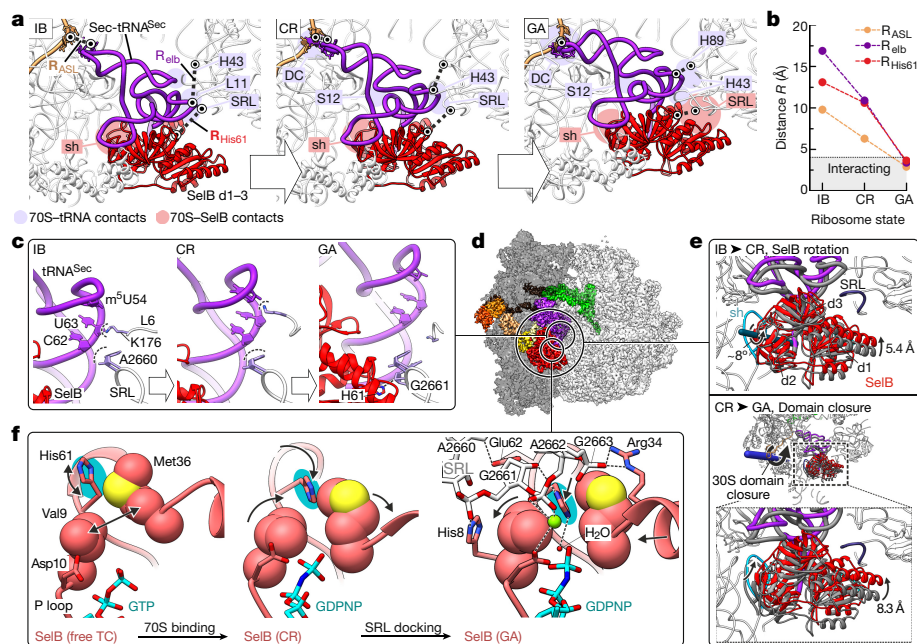


Figure 4 | Sequential docking of SelB–Sec-tRNA^{Sec} on the SRL.

a, Movements of the tRNA^{Sec} anticodon (R_{ASL}) and elbow (R_{elb}) and of SelB His61 (R_{His61}); sh, 30S shoulder (G357 to U368 of 16S rRNA); DC, decoding centre; H43 and H89, helices of 23S rRNA. **b**, State-specific distance changes. R_{ASL} (N3 of C35 in tRNA^{Sec} to N1 of G in UGA); R_{elb} (tRNA^{Sec} elbow, C5' of Ψ 55 in tRNA^{Sec} to O2' of A2473 in H89); R_{His61} (ND1 of His61 in SelB to O2' of G2661 in SRL). **c**, SRL interactions with Sec-tRNA^{Sec} and SelB His61. **d**, Overview of the ribosome–SelB complex in the GTPase-activated state indicating the areas for close-ups in **c**, **e**, **f**. **e**, Docking of SelB onto the SRL. Distances are measured at C α of Ala120.

Top, rotation of SelB relative to the 30S shoulder (sh) upon transition from initial binding to codon reading states. Bottom, rotation of SelB together with tRNA^{Sec} and the 30S shoulder upon 30S domain closure in the GTPase-activated state. **f**, The GTPase site. Left, free ternary complex. The two-headed arrows denote spontaneous fluctuations of SelB His61 (cyan) and of Val9 and Met36. Middle, inactive flipped-in conformation of His61 upon ribosome binding in the codon reading state, potentially stabilized by Val9. Right, active conformation of SelB docked on the SRL in the GTPase-activated state stabilized by the ribosome–SelB interaction network.

the complex. To understand the dynamics of those rearrangements, we carried out 24- μ s all-atom explicit-solvent molecular dynamics simulations of SelB–GTP–Sec-tRNA^{Sec} (Fig. 3c, Extended Data Fig. 6a, b; Methods). We mapped the free-energy conformational landscape and identified the lowest-energy structural model of the ternary complex free in solution. All tRNA^{Sec} conformations observed on the ribosome are sampled spontaneously within less than a microsecond in solution (Fig. 3c). Given that the GTPase activation takes milliseconds (Fig. 2c), the conformational changes of the tRNA, though clearly essential, do not seem to be rate-limiting for the GTPase activation. Of the large pool of potential tRNA^{Sec} structures accessible in solution, the ribosome selectively stabilizes specific subsets of conformations depending on the sum of all interactions in a given state (Fig. 3b, c), thereby rectifying the structural fluctuations of the complexes towards the GTPase-activated state.

Docking of SelB onto the SRL

The SRL constitutes the key activator of the GTPase activity for all translational GTPases³⁹. Notably, docking of the SelB GTP-binding domain on the SRL occurs only upon formation of the GTPase-activated state (Fig. 4a). Whereas the tRNA^{Sec} anticodon moves towards the codon in the 30S decoding centre, the elbow region changes its interactions from the initial contact with the SRL and protein L11 (initial binding state) to SRL and H43 of 23S rRNA (codon reading state) before docking onto H89 of 23S rRNA (Fig. 4a). The movements of the tRNA^{Sec} anticodon towards the codon and of the tRNA elbow towards H89, as well as the formation of the GTPase-activated state (monitored by the position of the key GTPase residue His61 in SelB) are coupled (Fig. 4b). These motions control the progression of SelB towards the GTPase-activated state, in line with the major role of tRNA in signalling codon recognition^{36–38}.

Stepwise tRNA^{Sec} docking is facilitated by several elements of the ribosome, including the L11–rRNA arm (consisting of protein L11 and

helices H43 and H44 of 23S rRNA), the SRL (together with protein L6), protein S12, and H69 of 23S rRNA that guide tRNA^{Sec} to the decoding centre. The L11–rRNA arm acts as a relay that hands over the tRNA during transitions from state to state (Extended Data Fig. 7a). Binding of tRNA^{Sec} results in a >10 Å displacement of the entire L11–rRNA in the initial binding state. In subsequent states, the L11–rRNA arm relaxes stepwise into its initial position. Notably, in the initial binding and codon reading states, the SRL interacts with tRNA^{Sec}, rather than with SelB (Fig. 4c). A2660 acts as a platform for docking of the T-stem residues of tRNA^{Sec}, but the contacts with tRNA^{Sec} change between the initial binding and codon reading states. Further movement of tRNA^{Sec} in the GTPase-activated state disrupts these contacts and enables docking of SelB onto the SRL. On the 30S subunit, protein S12 engages the acceptor stem of tRNA^{Sec} in the codon reading state and serves as a pivot for tRNA^{Sec} rotation into the GTPase-activated state (Extended Data Fig. 7b). A1913 of H69, in turn, reaches into the decoding centre and may help to guide the anticodon towards the codon in the aminoacyl site during the transition from the codon reading to the GTPase-activated state by interacting with the functionally important modified residue isopentenyl-A37 in tRNA^{Sec} (ref. 40; Extended Data Fig. 5a).

In the initial binding, codon reading and GTPase-activated states, SelB maintains its interactions with tRNA^{Sec} and follows its movements (Fig. 4a, e). SelB domain 2 remains bound to the shoulder of 16S rRNA, while adapting to the changes in the position of tRNA^{Sec}. Upon transition from the initial binding to the codon reading state, this adaptation requires a rotation of domain 2 relative to the 30S subunit shoulder, which changes the interactions and moves SelB slightly towards the 50S subunit (Fig. 4e, upper panel, and Extended Data Fig. 7c). In the GTPase-activated state, as a consequence of codon recognition and the global 30S domain closure, SelB rotates together with tRNA^{Sec} and the 30S shoulder. Domain 1 of SelB—comprising several charged residues (Arg34, Glu62 and Arg116)—moves into close proximity to

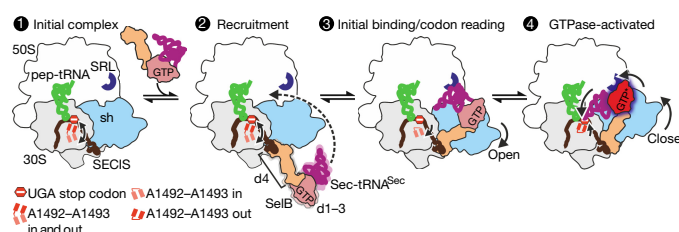


Figure 5 | Mechanism of ribosome-dependent GTPase activation of SelB-Sec-tRNA^{Sec} upon UGA recoding by Sec. See text for explanation. pep-tRNA, peptidyl-tRNA; d, domains of SelB; dark red glow, GTPase activation of SelB; arrows 'open' and 'close', domain opening and closure of the 30S shoulder (sh).

the SRL (Fig. 4e, bottom panel). These residues may help to promote the small interdomain motion of SelB that is required to completely dock onto the SRL (Extended Data Fig. 7d). The molecular dynamics simulations of the free SelB-GTP-Sec-tRNA^{Sec} complex indicate that this interdomain motion in SelB is rapidly sampled in solution (Extended Data Fig. 7e) and largely independent of tRNA dynamics (Extended Data Fig. 6c).

GTPase activation

The active site of all translational GTPases is highly conserved. The invariant histidine (His61 in SelB, His84 in EF-Tu) is essential for rapid GTP hydrolysis in all translational GTPases^{41–43}. The molecular dynamics simulations of SelB-GTP-Sec-tRNA^{Sec} in solution suggest that His61 can rapidly fluctuate between a flipped-out (pointing away from the γ -phosphate) and a flipped-in (reaching towards the γ -phosphate) conformation (Fig. 4f, Extended Data Fig. 7f, g), similarly to His84 in EF-Tu^{44,45}. In the codon reading state the conformational equilibrium of His61 is shifted towards a partially flipped-in conformation (Fig. 4f). The bulky residue Val9 of SelB appears to stabilize this inactive conformation, thereby impeding further movement of His61 towards the nucleotide (Fig. 4f), which explains the low rate of spontaneous GTP hydrolysis before codon recognition (Fig. 2c).

In the GTPase-activated state, SelB adopts a conformation that is similar to that of the pre-hydrolysis state of EF-Tu^{35,38} (Fig. 4f, Extended Data Fig. 7h). The universally conserved residue G2661 of the SRL stabilizes His61 of SelB in the flipped-in conformation, pointing towards the water molecule aligned for the attack on the γ -phosphate. Val9 stacks onto the His61 imidazole ring, providing additional stabilization of the active conformer. The phosphate of A2662 coordinates a Mg²⁺ ion, which may be important in positioning Asp10 in SelB (homologous to Asp21 in EF-Tu, which is crucial for GTP hydrolysis⁴²); a similar Mg²⁺ coordination is seen with the corresponding residue Asp22 of elongation factor G in the pre-hydrolysis state on the ribosome⁴⁶. Further SelB residues (His8, Arg34, Glu62, Arg116) stabilize the interactions with the SRL (Extended Data Fig. 7d, right panel). Given the evolutionary conservation of the residues constituting the GTPase centre, GTP hydrolysis is likely to follow the same universal pathway for all translational GTPases.

Conclusions

The present data show how SelB delivers Sec-tRNA^{Sec} to the ribosome, provide insights into the local and global conformations of the ribosome upon codon recognition by Sec-tRNA^{Sec} and suggest the mechanism for GTPase activation of SelB (Fig. 5). Sec-tRNA^{Sec} and SelB in solution can sample many different conformations. The interactions with the ribosome rectify these spontaneous fluctuations and guide the progression towards GTPase activation. In the initial complex, the SECIS element in the mRNA is exposed for SelB binding and the universally conserved bases A1492 and A1493 in the decoding centre fluctuate between flipped-in and flipped-out conformations (Fig. 5, step 1). Recruitment to SECIS tethers SelB to the vicinity of the ribosome; this contact

between SelB domain 4 and the SECIS is maintained in all subsequent steps (Fig. 5, step 2). Initial binding of SelB-Sec-tRNA^{Sec} to the ribosome induces an open conformation of the 30S subunit and stabilizes A1492 and A1493 in a flipped-in conformation, thereby facilitating codon reading (Fig. 5, step 3). Subsequent codon recognition triggers local closure of the decoding site, with A1492 and A1493 flipped out, and a global domain closure of the 30S subunit. The closed conformation facilitates re-positioning of the tRNA and docking of SelB on the SRL. Docking at the SRL aligns key residues at the nucleotide binding pocket of SelB and results in GTPase activation (Fig. 5, step 4). The structure of the GTPase-activated state is likely to be universal for translational GTPases in all domains of life. The docking of GTPases onto the SRL as a result of correct codon-anticodon complex formation and 30S domain closure may represent a common mechanism by which the ribosome ensures the pre-hydrolysis selection of the cognate aa-tRNA delivered by SelB, EF-Tu, or their eukaryotic homologues. This work emphasizes the power of the combination of high-resolution cryo-EM, rapid kinetics and molecular dynamics simulations for the reconstruction of multistep reaction pathways and understanding the function of dynamic molecular machines.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 July; accepted 24 October 2016.

Published online 14 November 2016.

- Böck, A. Biosynthesis of selenoproteins—an overview. *Biofactors* **11**, 77–78 (2000).
- Lescure, A., Fagegaltier, D., Carbon, P. & Krol, A. Protein factors mediating selenoprotein synthesis. *Curr. Protein Pept. Sci.* **3**, 143–151 (2002).
- Berchtold, H. *et al.* Crystal structure of active elongation factor Tu reveals major domain rearrangements. *Nature* **365**, 126–132 (1993).
- Paleskava, A., Konevega, A. L. & Rodnina, M. V. Thermodynamic and kinetic framework of selenocysteyl-tRNA^{Sec} recognition by elongation factor SelB. *J. Biol. Chem.* **285**, 3014–3020 (2010).
- Forchhammer, K., Boesmler, K. & Böck, A. The function of selenocysteine synthase and SELB in the synthesis and incorporation of selenocysteine. *Biochimie* **73**, 1481–1486 (1991).
- Itoh, Y., Sekine, S., Suetsugu, S. & Yokoyama, S. Tertiary structure of bacterial selenocysteine tRNA. *Nucleic Acids Res.* **41**, 6729–6738 (2013).
- Zinoni, F., Heider, J. & Böck, A. Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *Proc. Natl Acad. Sci. USA* **87**, 4660–4664 (1990).
- Itoh, Y., Sekine, S. & Yokoyama, S. Crystal structure of the full-length bacterial selenocysteine-specific elongation factor SelB. *Nucleic Acids Res.* **43**, 9028–9038 (2015).
- Leibundgut, M., Frick, C., Thanbichler, M., Böck, A. & Ban, N. Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors. *EMBO J.* **24**, 11–22 (2005).
- Kromayer, M., Witting, R., Tormay, P. & Böck, A. Domain structure of the prokaryotic selenocysteine-specific elongation factor SelB. *J. Mol. Biol.* **262**, 413–420 (1996).
- Nissen, P., Thirup, S., Kjeldgaard, M. & Nyborg, J. The crystal structure of Cys-tRNA^{Cys}-EF-Tu-GDPNP reveals general and specific features in the ternary complex and in tRNA. *Structure* **7**, 143–156 (1999).
- Hatfield, D. L., Carlson, B. A., Xu, X. M., Mix, H. & Gladyshev, V. N. Selenocysteine incorporation machinery and the role of selenoproteins in development and health. *Prog. Nucleic Acid Res. Mol. Biol.* **81**, 97–142 (2006).
- Metanis, N. & Hilvert, D. Natural and synthetic selenoproteins. *Curr. Opin. Chem. Biol.* **22**, 27–34 (2014).
- Bulteau, A. L. & Chavatte, L. Update on selenoprotein biosynthesis. *Antioxid. Redox Signal.* **23**, 775–794 (2015).
- Ringquist, S. *et al.* Recognition of the mRNA selenocysteine insertion sequence by the specialized translational elongation factor SELB. *Genes Dev.* **8**, 376–385 (1994).
- Moazed, D. & Noller, H. F. Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* **57**, 585–597 (1989).
- Förster, C., Ott, G., Forchhammer, K. & Sprinzl, M. Interaction of a selenocysteine-incorporating tRNA with elongation factor Tu from *E. coli*. *Nucleic Acids Res.* **18**, 487–491 (1990).
- Rudinger, J., Hillenbrandt, R., Sprinzl, M. & Giegé, R. Antideterminants present in mini-helix(Sec) hinder its recognition by prokaryotic elongation factor Tu. *EMBO J.* **15**, 650–657 (1996).
- Yikilmaz, E., Chapman, S. J., Schrader, J. M. & Uhlenbeck, O. C. The interface between *Escherichia coli* elongation factor Tu and aminoacyl-tRNA. *Biochemistry* **53**, 5710–5720 (2014).

20. Schrader, J. M., Chapman, S. J. & Uhlenbeck, O. C. Understanding the sequence specificity of tRNA binding to elongation factor Tu using tRNA mutagenesis. *J. Mol. Biol.* **386**, 1255–1264 (2009).
21. Asahara, H. & Uhlenbeck, O. C. The tRNA specificity of *Thermus thermophilus* EF-Tu. *Proc. Natl Acad. Sci. USA* **99**, 3499–3504 (2002).
22. Diaconu, M. *et al.* Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121**, 991–1004 (2005).
23. Maracci, C. & Rodnina, M. V. Review: Translational GTPases. *Biopolymers* **105**, 463–475 (2016).
24. Pape, T., Wintermeyer, W. & Rodnina, M. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J.* **18**, 3800–3807 (1999).
25. Heider, J., Baron, C. & Böck, A. Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. *EMBO J.* **11**, 3759–3766 (1992).
26. Kotini, S. B., Peske, F. & Rodnina, M. V. Partitioning between recoding and termination at a stop codon-selenocysteine insertion sequence. *Nucleic Acids Res.* **43**, 6426–6438 (2015).
27. Wohlgemuth, I., Pohl, C., Mittelstaet, J., Konevega, A. L. & Rodnina, M. V. Evolutionary optimization of speed and accuracy of decoding on the ribosome. *Phil. Trans. R. Soc. Lond. B* **366**, 2979–2986 (2011).
28. Marshall, R. A., Aitken, C. E., Dorywalska, M. & Puglisi, J. D. Translation at the single-molecule level. *Annu. Rev. Biochem.* **77**, 177–203 (2008).
29. Thompson, R. C., Dix, D. B. & Karim, A. M. The reaction of ribosomes with elongation factor Tu.GTP complexes. Aminoacyl-tRNA-independent reactions in the elongation cycle determine the accuracy of protein synthesis. *J. Biol. Chem.* **261**, 4868–4874 (1986).
30. Ruusala, T., Ehrenberg, M. & Kurland, C. G. Is there proofreading during polypeptide synthesis? *EMBO J.* **1**, 741–745 (1982).
31. Ogle, J. M., Murphy, F. V., Tarry, M. J. & Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111**, 721–732 (2002).
32. Ogle, J. M. *et al.* Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902 (2001).
33. Dunkle, J. A. *et al.* Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332**, 981–984 (2011).
34. Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329–333 (2010).
35. Voorhees, R. M., Schmeing, T. M., Kelley, A. C. & Ramakrishnan, V. The mechanism for activation of GTP hydrolysis on the ribosome. *Science* **330**, 835–838 (2010).
36. Piepenburg, O. *et al.* Intact aminoacyl-tRNA is required to trigger GTP hydrolysis by elongation factor Tu on the ribosome. *Biochemistry* **39**, 1734–1738 (2000).
37. Cochella, L. & Green, R. An active role for tRNA in decoding beyond codon:anticodon pairing. *Science* **308**, 1178–1180 (2005).
38. Schmeing, T. M., Voorhees, R. M., Kelley, A. C. & Ramakrishnan, V. How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat. Struct. Mol. Biol.* **18**, 432–436 (2011).
39. Moazed, D., Robertson, J. M. & Noller, H. F. Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA. *Nature* **334**, 362–364 (1988).
40. Warner, G. J. *et al.* Inhibition of selenoprotein synthesis by selenocysteine tRNA[Ser]^{Sec} lacking isopentenyladenosine. *J. Biol. Chem.* **275**, 28110–28119 (2000).
41. Daviter, T., Wieden, H. J. & Rodnina, M. V. Essential role of histidine 84 in elongation factor Tu for the chemical step of GTP hydrolysis on the ribosome. *J. Mol. Biol.* **332**, 689–699 (2003).
42. Maracci, C., Peske, F., Dannies, E., Pohl, C. & Rodnina, M. V. Ribosome-induced tuning of GTP hydrolysis by a translational GTPase. *Proc. Natl Acad. Sci. USA* **111**, 14418–14423 (2014).
43. Cunha, C. E. *et al.* Dual use of GTP hydrolysis by elongation factor G on the ribosome. *Translation (Austin)* **1**, e24315 (2013).
44. Adamczyk, A. J. & Warshel, A. Converting structural information into an allosteric-energy-based picture for elongation factor Tu activation by the ribosome. *Proc. Natl Acad. Sci. USA* **108**, 9827–9832 (2011).
45. Wallin, G., Kamerlin, S. C. & Aqvist, J. Energetics of activation of GTP hydrolysis on the ribosome. *Nat. Commun.* **4**, 1733 (2013).
46. Tourigny, D. S., Fernández, I. S., Kelley, A. C. & Ramakrishnan, V. Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science* **340**, 1235490 (2013).

Acknowledgements We thank F. Würriehausen, A. Bursy, O. Geintzer, S. Kappler, C. Kothe, T. Niese, T. Wiles, and M. Zimmermann for expert technical assistance and M. Lüttich and T. Koske for support in high-performance computation. The work was supported by the Deutsche Forschungsgemeinschaft Grant FOR 1805 (to H.S., M.V.R. and H.G.) and by the Sonderforschungsbereich 860 (to R.F.).

Author Contributions N.F. conceived the project, performed cryo-EM experiments and data analysis and drafted the paper. P.N. built atomic models and performed pseudo-crystallographic refinement. L.V.B. performed and analyzed molecular dynamics simulations. C.M. performed the kinetic analysis. A.P. and A.L.K. prepared ribosome complexes for cryo-EM. Z.W. and G.F.S. performed homology modelling. N.F., M.V.R. and H.S. finalized the paper with inputs from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.F. (niels.fischer@mpiibpc.mpg.de) and H.S. (hstark1@gwdg.de).

Reviewer Information *Nature* thanks A. Amunts, M. Wahl and S. Yokoyama for their contribution to the peer review of this work.

METHODS

Ribosomes, factors and RNAs. Ribosomes from *E. coli* MRE 600, initiation factors (IF1, IF2, IF3), and fMet-tRNA^{fMet} were prepared as described^{47–49}. *E. coli* SelA and SelD (gift from M. Wahl, Free University of Berlin) and SelB⁵⁰ containing a hexahistidine tag (gift from A. Böck, LMU Munich) were expressed in BL21(DE3) cells and purified according to published protocols^{4,51}. Biochemical analysis showed that SelB carrying the His-tag is fully functional in the interactions with guanine nucleotides, SECIS elements, and Sec-tRNA^{Sec} and in mediating UGA recoding *in vivo*^{4,50}. tRNA^{Sec} was overexpressed in *E. coli* BL21 transformed with the plasmid pCB2013 (gift from A. Böck, LMU Munich)⁵² and purified and aminoacylated as described^{4,53,54}. The mRNA UGA was a derivative of mL755⁵⁴ (Extended Data Fig. 4d). The templates for mRNAs UUC and iSECIS (Extended Data Fig. 4e, f) were prepared by PCR mutagenesis. Transcription and refolding of the mRNAs were performed as described^{54,55}.

Complex preparation for cryo-EM. 70S ribosomes (3 μ M) were incubated with IF1, IF2, IF3 (4.5 μ M), mRNA (15 μ M), and f[³H]Met-tRNA^{fMet} (7 μ M) in buffer A (50 mM Tris-HCl (pH 7.5), 70 mM NH₄Cl, 30 mM KCl, 7 mM MgCl₂, 2 mM DTT) with 1 mM GTP for 30 min at 37 °C. Initiation efficiency was close to 100% as verified by nitrocellulose binding. Initiation complexes were purified by gel filtration on a Biosuite 450 HR 5 μ m column (Waters). To prepare the ribosome–SelB complexes, ternary complexes SelB–GDPNP–Sec-tRNA^{Sec} were prepared in buffer B (50 mM Hepes-KOH, pH 7.5, 70 mM NH₄Cl, 30 mM KCl, 7 mM MgCl₂, 2 mM DTT) by incubating SelB (1 μ M) with GDPNP (2 mM) for 4 min at 37 °C, adding Sec-tRNA^{Sec} (1 μ M) and incubating for 2 min at 23 °C. Ternary complex (0.5 μ M) was incubated with initiation complex (0.06 μ M) at 0 °C in buffer B supplemented with 0.6 mM spermine and 0.4 mM spermidine before application onto EM grids.

GTPase assay. Initiation complexes were prepared as described⁴². Ternary complex SelB–[γ -³²P]GTP–Sec-tRNA^{Sec} was prepared in buffer A, by incubating SelB (9 μ M) with [γ -³²P]GTP (60 μ M) for 20 min at room temperature (RT), followed by addition of Sec-tRNA^{Sec} (9 μ M) and incubation for 2 min at room temperature. Unbound [γ -³²P]GTP was removed by gel filtration⁵⁶. The GTPase activity of SelB was determined at single round conditions, by mixing purified ternary complex (0.05 μ M) with UGA, iSECIS and UUC mRNA-programmed initiation complexes at the indicated concentrations. Intrinsic GTP hydrolysis was measured in the absence of ribosomes. When necessary, quench-flow experiments were performed in a KinTek apparatus. Reactions were quenched with formic acid (25% v/v) and the extent of GTP cleavage was determined by thin layer chromatography and phosphor imaging⁵⁷. The rate of GTP hydrolysis was determined by exponential fitting of the time courses using GraphPad Prism software (GraphPad Software, Inc.); time courses were normalized to the respective reaction end levels.

Cryo-EM analysis. Cryo-EM grids were prepared by applying 5 μ l of initiation complex–SelB–GDPNP–Sec-tRNA^{Sec} complexes onto EM grids (Quantifoil 3.5/1 μ m, Jena) covered with pre-floated continuous carbon and subsequently vitrified using a Vitrobot Mark IV (FEI Company, Eindhoven) operated at 4 °C and 100% humidity. 4,000 \times 4,000 images (12,681 in total) were acquired in the integration mode by spot-scanning (3 \times 3 images per 3.5 μ m hole in the Quantifoil carbon film) on a Falcon 2 direct detector (FEI Eindhoven) using a Titan Krios microscope (FEI Eindhoven), fitted with a XFEG electron source (FEI Eindhoven) and a spherical aberration (Cs)-corrector (CEOS Heidelberg). Acquisition was at 300 kV acceleration voltage, an electron dose of \sim 30 \pm 5 electrons per \AA^2 (determination based on calibration with a Faraday cup), -0.7 to -2.6 μ m defocus and a nominal magnification of 118,000 \times resulting in a final pixel size of \sim 1.16 \AA (determined by optimizing correlation of the final 3D map with atomic coordinates from X-ray crystallography⁵⁸). The Cs-corrector was tuned as described⁵⁸ to reduce electron-optical aberrations, linear distortion (to $<0.1\%$) and axial coma (usimageshift tuning) resulting from the spot-scanning procedure. Ribosome particle images were selected and corrected locally for the contrast-transfer function as described⁵⁸. Good quality particle images were selected according to power spectra quality (that is, showing Thon rings better than 3.5 \AA), yielding a total of 969,526 ribosome particles. An initial 3D structure was computed from randomly selected 100,000 particles to judge the compositional and conformational homogeneity of the image dataset. Even at low-resolution (\sim 10 \AA) this average structure showed only very weak, scattered density in the factor-binding region, which disappeared completely upon further refinement to higher-resolution (\sim 3.7 \AA) (Extended Data Fig. 1c). This observation indicated that the occupancy of ribosome particles with SelB–Sec-tRNA^{Sec} was low, in line with the low Sec incorporation efficiency *in vivo* and *in vitro*²⁶.

The dataset was sorted computationally in a hierarchical manner³⁴ (Extended Data Fig. 1a, b). First, images were sorted according to global ribosome conformation, as described⁵⁸ (step 1), which allowed us to discard low quality particles

and 50S subunits. Subsequently (step 2), particles were sorted according to ligand occupancy using supervised classification by projection matching⁵⁹ on the basis of a library of ribosome–ligand cryo-EM maps^{34,60}. Importantly, this library entailed a cryo-EM map of the canonical ribosome–EF–Tu complex⁶⁰ as potential reference for the ribosome–SelB complex to avoid any reference bias for SelB–Sec-tRNA^{Sec}, for which no structure was available. To further avoid any high-resolution reference bias, sorting in both steps was performed using low-pass filtered reference maps and particle images binned to about 6.9 \AA per pixel. The resulting maps of ribosome–ligand complexes showed all specific features expected for the Sec system (that is, the SECIS in the mRNA, the extra-long variable arm of tRNA^{Sec}, and domain 4 of SelB). In step 3, the resulting populations of ribosome–ligand particles were further rectified from low-quality particles by three-dimensional (3D) classification in Relion 1.2 and 1.3 (ref. 61). Focused 3D classification in Relion 1.3 without alignment was used to isolate three particle sub-populations that show distinct states of SelB–Sec-tRNA^{Sec} on the ribosome (step 4); particle images were coarsened to 3.16 \AA per pixel for this step to improve the speed and robustness of classification. Finally, to further enrich for ribosome particles of good quality containing SelB within these sub-populations, supervised classification was used in step 5 to assign particles either to the respective ribosome–SelB complex or ribosomes with only peptidyl-site tRNA, but not SelB–Sec-tRNA^{Sec} bound; again a pixel sampling of 3.16 \AA and low-pass filtered reference maps were used to avoid any high-resolution reference bias. In a similar manner, the particle populations of ribosomes carrying only the peptidyl-site tRNA or two tRNAs in the classical state were further computationally cleaned up by supervised classification. The six homogenous ribosome–ligand particle populations resulting from this sorting procedure were refined to high resolution according to the gold-standard procedure in Relion 1.3 and overall resolution of the final maps was determined using high-resolution noise substitution⁶² in Relion 1.4 (Extended Data Fig. 2a). The final cryo-EM maps were sharpened as described⁵⁸; and for each reconstruction local, resolution maps were computed using Resmap⁶³ (Extended Data Fig. 2f, g). If not denoted otherwise, densities in the figures are rendered at 2σ , except winged-helix motifs 3–4 and SECIS of the GTPase-activated state, which are rendered at 1.3–1.7 σ . Figures were prepared using UCSF Chimera⁶⁴.

Atomic model refinement. Pseudo-crystallographic refinement and model building. For initial model building and refinement, the cryo-EM density map of the GTPase-activated state was sharpened by applying a B factor of -120\AA^2 , filtered to the estimated highest resolution limit and masked using a pseudo bulk solvent envelope. The mask was obtained by merging the cryo-EM map filtered at different frequencies (9.0 \AA , 6.0 \AA and 4.0 \AA) using the RAVE package⁶⁵ and UCSF Chimera⁶⁴. The masked cryo-EM density map was converted to reciprocal space structure factors using Crystallography and NMR System (CNS)^{66,67} without employing phase significance blurring scale factors derived from FSC values.

Initial rigid body fit of an atomic model of the kirromycin-stalled *E. coli* ribosome complex⁵⁸ lacking EF–Tu and Phe-tRNA^{Phe} was performed against the masked cryo-EM map of the GTPase-activated state using Chimera, followed by rigid body refinement of individual chains in PHENIX⁶⁸ program. The atomic model of SelB with exception of tandem winged-helix motifs 3 and 4 was obtained by combining manual rebuilding and homology modelling with density-guided energy optimization, as implemented in the Rosetta package^{69,70}, employing a template derived from *Methanococcus maripaludis* SelB (PDB ID: 4AC9)⁹ and alignment provided by the HHPRED server⁷¹. SelB tandem winged-helix motifs (wh) 1 and 2 were modelled based on the crystal structure of the isolated domain 4 of SelB from *Moorella thermoacetica* (PDB ID: 1LVA)⁷²; and wh3 and wh4 with mRNA SECIS were modelled on the basis of the crystal structure of *E. coli* SECIS RNA bound to the domain of elongation factor SelB (PDB ID: 2PJP)⁷³ using Rosetta and Coot⁷⁴. Sec-tRNA^{Sec} was manually modelled on the basis of the crystal structure of tRNA^{Sec} in complex with seryl-tRNA synthetase (PDB ID: 3W3S)⁶ with Coot and Rcrane⁷⁵ and optimized by ERRASER⁷⁶. The catalytic water molecule in the GTPase active centre of SelB was modelled on the basis of the superposition from PDB ID 4V5L³⁵ into available density. Before global optimization of the atomic model of ribosome was performed, models of all individual protein chains were relaxed against the masked cryo-EM map filtered to 3.8 \AA resolution (GTPase-activated state) using Rosetta. The assembled complete ribosome model was further minimized in both real and reciprocal space (ERRASER, phenix.real_space_refine⁷⁷, phenix.refine) with alternating cycles of manual rebuilding in Coot and monitoring the local fit to the density with RESOLVE⁷⁸. Refinement progress was additionally monitored by calculating the Pearson correlation coefficient (CC_{work}; ref. 79), as well as the Fourier shell correlation between the cryo-EM reconstruction and a model map (FSC_{work}). Both real and reciprocal space refinements in PHENIX employed automatically generated restraints and additional pseudo DEN (Deformable Elastic Network) restraints in order to maintain chemically important interactions. Group atomic displacement factors (ADP) were refined exclusively during reciprocal space refinement steps. During real space

refinement steps in phenix.real_space_refine Ramachandran plot restraints were enabled. For parts of the model exhibiting larger conformational differences and/or lower local map resolution, additional cycles of real space refinement and manual fitting were performed against experimental map filtered to lower resolution, which was gradually increased between subsequent refinement steps until convergence. To maintain the intermolecular interactions of selected model fragments within local environment, all residues within at least 15 Å radius were included. For final refinement steps the cryo-EM map was sharpened⁵⁸ and masked using a smoothed model-based envelope generated within 3.0 Å radius around atoms of the ribosome model of the GTPase-activated state before conversion to reciprocal space structure factors; for these steps and improved visualization the cryo-EM maps resolved at ≤ 3.6 Å resolution (initial complex, classical and GTPase-activated states) were resampled to a pixel size of 0.789 Å. The final model of the GTPase-activated state consisting of 154,136 individual atoms was refined to 24.90% and 0.936 for R_{work} and CC_{work} (definition is given below), respectively. The model of the GTPase-activated state exhibits a good stereochemistry with 90.87% of residues in the most favoured region and 0.88% residues in the disallowed region of the Ramachandran plot, protein side chain outliers of 2.83% and all atom clash score 15.09.

The remaining states (initial complex, initial binding, codon reading, classical and hybrid states) were modelled based on the final model of GTPase-activated state in an analogous way. For cryo-EM reconstructions resolved at a resolution lower than 3.6 Å, a 3.8 Å radius for generating a smooth model-based envelope was chosen before generating reciprocal space structure factors used for the final refinement steps. F_{SCwork} and $CC_{\text{work/overall}}$ were calculated in a resolution-dependent manner using SFALL program (CCP4 suite⁶⁰). Real space correlation coefficients (RSCC) were calculated using RESOLVE. All atomic models fit well the experimental cryo-EM map as judged based on the following criteria. (1) The Pearson correlation coefficient (CC_{work}) calculated between model and map structure factors (F_{model} and F_{EM}) used for refinement is greater than 0.2 for the highest resolution shell and the overall correlation coefficient (CC_{overall}) is not lower than 0.9. (2) The calculated F_{SCwork} value between model map coefficients (F_{model} , $\text{phase}_{\text{model}}$) and structure factors derived from the cryo-EM map (F_{EM} , phase_{EM}) used for refinement are not lower than 0.5 for the highest resolution shell and overall greater than 0.9. (3) The cumulative RSCC values are greater than 0.8, 0.6 and 0.4 for 53–75%, 88–94% and 95–97% of the residues, respectively. Detailed refinement statistics are presented in Extended Data Table 1, FSC and CC_{work} curves in Extended Data Fig. 2a–c.

Modelling of the recruitment complex. To build the initial recruitment complex, SelB domain 4 from the structure of the isolated SECIS-domain 4 complex (wh2 to wh4; PDB ID: 2PLY)⁷³ was docked onto the initial complex by superimposing the SECIS. Then, wh2 was used to model our structure of SelB–GDPNP–Sec-tRNA^{Sec} onto the initial complex-domain 4 complex.

Refinement against half maps. Refinements of final models against datasets obtained from two half maps were performed at 2.5 Å resolution in PHENIX using several cycles of real space refinement followed by reciprocal space refinement employing reference model restraints. To remove possible model bias a random shift of 0.3 Å was applied to all atomic positions before real space refinement. The FSC and CC were calculated between the models and the masked half-maps used for refinement, as well as between the model and the other half-map for cross-validation. The individual half-maps were masked using a smoothed mask derived from the respective refined model using a radius of 3.0 or 3.8 Å depending on the estimated highest resolution limit.

Molecular dynamics simulations. Simulation setup. To obtain the energetics and dynamics of the free ternary complex, molecular dynamics simulations were started from the ribosome-bound conformations (initial binding, codon reading, and GTPase-activated states). Coordinates of Sec-tRNA^{Sec}, SelB (amino acids 1–401), and GTP were extracted from the cryo-EM structures along with resolved water molecules and ions in the vicinity (< 5 Å). The system was protonated, solvated, and ions were added as described earlier⁸¹.

All simulations were carried out with Gromacs 5 (ref. 82) using the amberff12sb force field⁸³, and the SPC/E water model⁸⁴. Parameters for potassium and chloride ions were taken from ref. 85 and for modified nucleotides from Aduri *et al.*⁸⁶. Atom types for selenocysteine were obtained with ANTECHAMBER⁸⁷ and partial charges were determined using DFT-B3LYP with a 6–31/G* basis set. Bond and virtual-site constraints, temperature and pressure coupling were applied as described⁸¹.

For each of the three starting structures, the system was pre-equilibrated as described⁸¹ with potential-energy minimization and 50-ns molecular dynamics simulations with position restraints followed by release of position restraints during 20 ns. Production runs started at 70 ns. At times 170, 270, and 370 ns, coordinates were extracted from the trajectory, new velocities were assigned and new simulations were started (12 simulations; total production run simulation time of 24 μ s).

Conformational dynamics of the ternary complex. To address the question of whether intrinsic conformational changes are rate-limiting for the transitions between the free ternary complex and the ribosome-bound states, we carried out a principal component analysis (PCA)⁸⁸ on three subsets of atoms: ‘v-arm’ (P-atoms of variable arm), ‘no-v-arm’ (remaining peptidyl-atoms of tRNA), and ‘D1’ (C α -atoms of SelB domain D1). The trajectories were rigid-body fitted using T-stem and acceptor-stem peptidyl-atoms for the ‘v-arm’ and ‘no-v-arm’ subsets and D2 C α -atoms for the ‘D1’ subsets ($\Delta t = 10$ ps, discarding first 200 ns). The trajectories were concatenated and the atomic-displacement covariance-matrix was calculated. The first eigenvectors, sorted according to their eigenvalues, represent the most dominant conformational modes. For ‘no-v-arm’, the first two eigenvectors largely consist of a bending motion of the anticodon arm. To estimate the free-energy landscape of this motion, the projections of all frames onto these eigenvectors were binned. The relative free energy of each bin was calculated via $\Delta G_{\text{bin}} = -k_{\text{B}} T \ln(c_{\text{bin}}/c_{\text{max}})$, where k_{B} is the Boltzmann constant, T the temperature, c_{bin} the number of frames in the bin and c_{max} is the maximum of all c_{bin} (Fig. 4c, Extended Data Figs 6a, 7e).

For ‘no-v-arm’ conformational modes 1 and 2, the bending angles (α and β) as a function of the projection onto the respective mode were calculated. The vector between the centres of mass of U34–A36 peptidyl-atoms and of T-stem peptidyl-atoms was calculated. For each projection, the angle between the vector and the vector corresponding to the bin with $\Delta G = 0$ was calculated.

To monitor the local geometry of the active site in the free ternary complex, the minimal distances between His61 imidazole atoms and GTP atoms (R_{GTP}) as well as the van-der-Waals distance between Val9 and Met36 (R_{Gate}) were calculated ($\Delta t = 10$ ps, discarding first 200 ns). The resulting free-energy landscape is shown in Extended Data Fig. 7f.

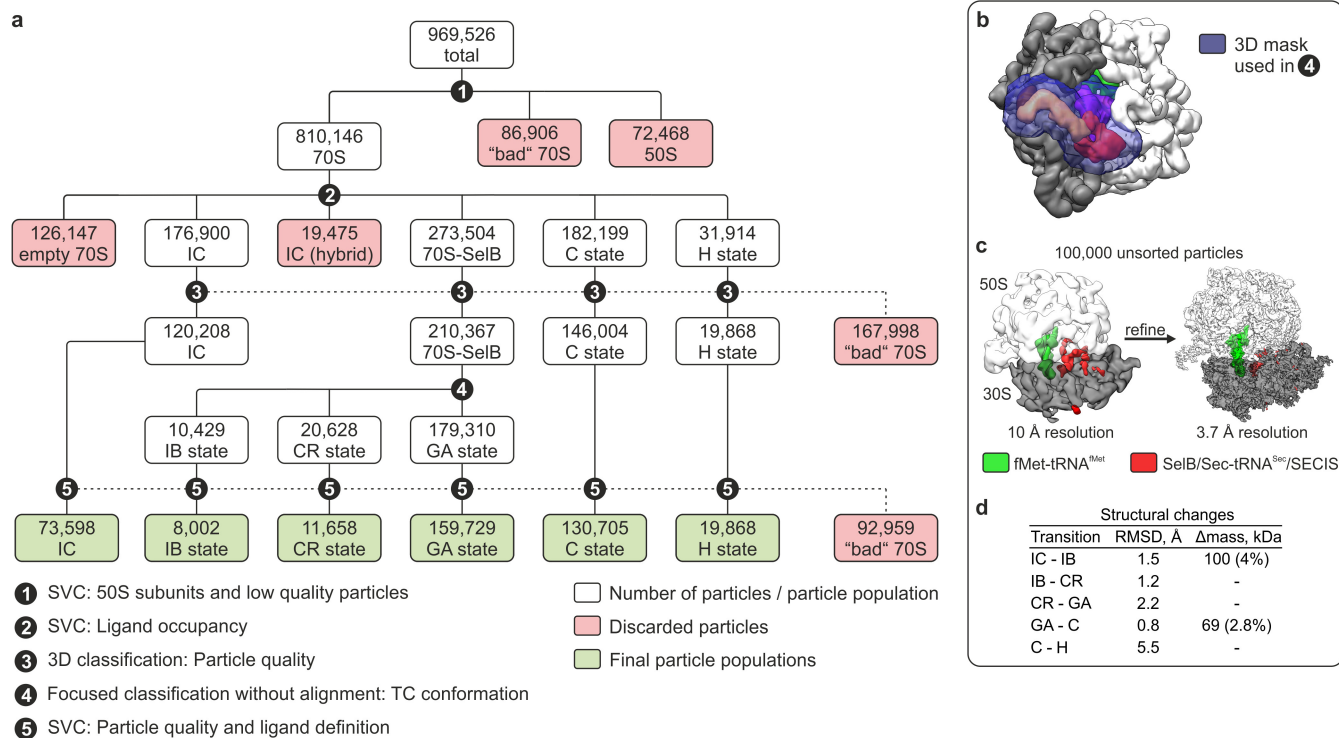
Transition rates for tRNA movements. To estimate the magnitudes of rates for transitions of tRNA^{Sec} in the free ternary complex between the conformations corresponding to the ribosome-bound states, the projections onto the ‘no-v-arm’ conformational modes 1 and 2 were calculated, thus obtaining 2-dimensional trajectories. For each simulation the number of transitions n between the region around the free-energy minimum (all bins with $0 \leq \Delta G \leq 0.5 k_{\text{B}} T \text{ mol}^{-1}$) and the ribosome-bound conformations was counted. For each ribosome-bound tRNA conformation a region was defined by an ensemble of 10,000 structures generated from the cryo-EM coordinates and b-factors. For each peptidyl-atom and ensemble structure, the coordinate obtained from cryo-EM was shifted in a random direction by distance d drawn from a normal distribution $p(d)$. The normal distribution $p(d)$ with $\mu = 0$ and $\sigma = \text{rmsf}$ (calculated from the b -factor) was set to zero for $d < \sigma$ and $d > \sigma$. The resulting structures were projected onto the ‘no-v-arm’ conformational modes. The region was defined as all bins with entries from these structures. The transition rates were calculated for each simulation by dividing n by the simulation time t_{sim} . The mean rates and standard deviations were obtained by weighting the individual transition rates by the corresponding t_{sim} .

Coupling of tRNA and SelB conformational modes. To address a possible coupling of ‘v-arm’ and ‘no-v-arm’ conformational modes, we investigated whether the projections onto these modes are correlated as would be expected for coupled motions. For all four pairs of eigenvectors v_a and v_b , where v_a is either ‘no-v-arm’ mode 1 or 2 and v_b is either ‘v-arm’ mode 1 or 2, the correlation coefficient of the projections was calculated. To estimate the statistical error, a bootstrapping method was applied. First, the autocorrelation time τ was defined by $f_{\text{ac}}(\tau) = \exp(-1)$, where $f_{\text{ac}}(\Delta t)$ is the autocorrelation function of the projection. For a given pair of eigenvectors v_a and v_b , the maximum of the two corresponding τ values (τ_{max}) was used as an interval to resample the projections, resulting in sets of $N = t_{\text{sim}}/\tau_{\text{max}}$ projections. This resampling was repeated 1,000 times and mean and standard deviations of the correlation coefficients were calculated for the simulations started from the individual ribosome-bound conformations and for all simulations combined (Extended Data Fig. 6b, red points). To check whether the resulting correlation coefficients are statistically significant, the standard deviations expected from uncorrelated projections were estimated by randomly drawing N projections (1,000 repetitions; Extended Data Fig. 6b, grey points).

To address coupling between tRNA dynamics and motions of SelB domain 1 (D1), the same analysis was carried out for ‘no-v-arm’ and ‘D1’ conformational modes (Extended Data Fig. 6c).

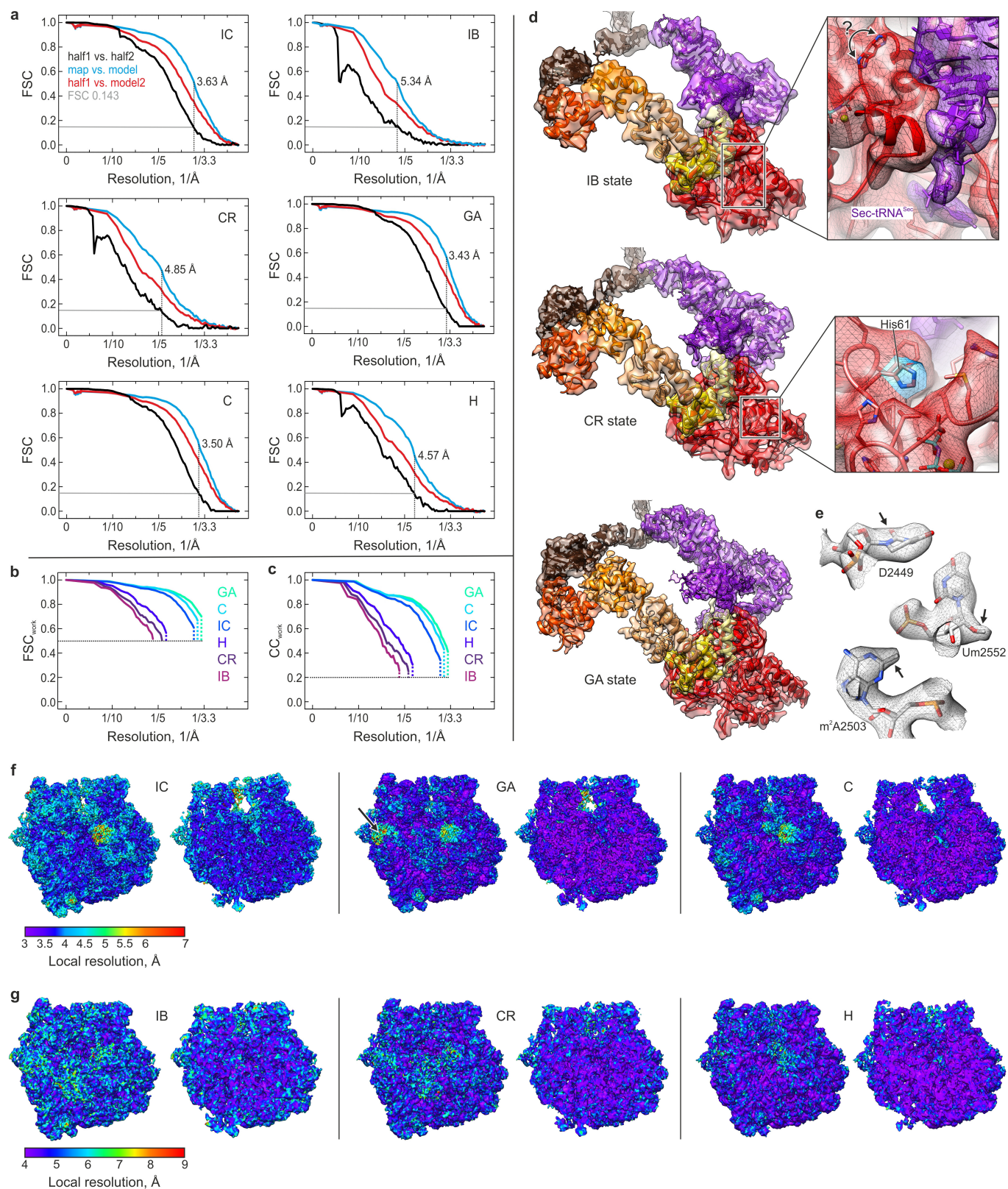
Data availability statement. Cryo-EM maps/coordinates of the atomic models for each state have been deposited in the Electron Microscopy Data Bank/Protein Data Bank (<http://www.emdatabank.org/>) with the following accession codes: initial complex, EMD-4121/5ZLA; initial binding state, EMD-4122/5ZLB; codon reading state, EMD-4123/5ZLC; GTPase-activated state, EMD-4124/5ZLD; classical state, EMD-4125/5ZLE; hybrid state, EMD-4126/5ZLF. Cryo-EM micrographs and particle images have been deposited in the EMPIAR database (<https://www.ebi.ac.uk/pdbe/emdb/empiar/>) with accession code EMPIAR-10077.

47. Rodnina, M. V. *et al.* Thiostrepton inhibits the turnover but not the GTPase of elongation factor G on the ribosome. *Proc. Natl Acad. Sci. USA* **96**, 9586–9590 (1999).
48. Milon, P. *et al.* Transient kinetics, fluorescence, and FRET in studies of initiation of translation in bacteria. *Methods Enzymol.* **430**, 1–30 (2007).
49. Rodnina, M. V. & Wintermeyer, W. GTP consumption of elongation factor Tu during translation of heteropolymeric mRNAs. *Proc. Natl Acad. Sci. USA* **92**, 1945–1949 (1995).
50. Thanbichler, M. & Böck, A. Purification and characterization of hexahistidine-tagged elongation factor SelB. *Protein Expr. Purif.* **31**, 265–270 (2003).
51. Paleskava, A., Konevega, A. L. & Rodnina, M. V. Thermodynamics of the GTP-GDP-operated conformational switch of selenocysteine-specific translation factor SelB. *J. Biol. Chem.* **287**, 27906–27912 (2012).
52. Thanbichler, M. & Böck, A. Selenoprotein biosynthesis: purification and assay of components involved in selenocysteine biosynthesis and insertion in *Escherichia coli*. *Methods Enzymol.* **347**, 3–16 (2002).
53. Kothe, U., Paleskava, A., Konevega, A. L. & Rodnina, M. V. Single-step purification of specific tRNAs by hydrophobic tagging. *Anal. Biochem.* **356**, 148–150 (2006).
54. Fischer, N. *et al.* Towards understanding selenocysteine incorporation into bacterial proteins. *Biol. Chem.* **388**, 1061–1067 (2007).
55. Konevega, A. L. *et al.* Purine bases at position 37 of tRNA stabilize codon-anticodon interaction in the ribosomal A site by stacking and Mg²⁺-dependent interactions. *RNA* **10**, 90–101 (2004).
56. Rodnina, M. V., Fricke, R. & Wintermeyer, W. Transient conformational states of aminoacyl-tRNA during ribosome binding catalyzed by elongation factor Tu. *Biochemistry* **33**, 12267–12275 (1994).
57. Gromadski, K. B. & Rodnina, M. V. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol. Cell* **13**, 191–200 (2004).
58. Fischer, N. *et al.* Structure of the *E. coli* ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* **520**, 567–570 (2015).
59. Valle, M. *et al.* Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. *EMBO J.* **21**, 3557–3567 (2002).
60. Stark, H. *et al.* Ribosome interactions of aminoacyl-tRNA and elongation factor Tu in the codon-recognition complex. *Nat. Struct. Biol.* **9**, 849–854 (2002).
61. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
62. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
63. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
64. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
65. Kleywegt, G. J. & Jones, T. A. Software for handling macromolecular envelopes. *Acta Crystallogr. D* **55**, 941–944 (1999).
66. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nat. Protocols* **2**, 2728–2733 (2007).
67. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
68. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
69. DiMaio, F. *et al.* Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature* **473**, 540–543 (2011).
70. DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. Refinement of protein structures into low-resolution density maps using rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).
71. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
72. Selmer, M. & Su, X. D. Crystal structure of an mRNA-binding fragment of *Morella thermoacetica* elongation factor SelB. *EMBO J.* **21**, 4145–4153 (2002).
73. Soler, N., Fourmy, D. & Yoshizawa, S. Structural insight into a molecular switch in tandem winged-helix motifs from elongation factor SelB. *J. Mol. Biol.* **370**, 728–741 (2007).
74. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
75. Keating, K. S. & Pyle, A. M. Semiautomated model building for RNA crystallography using a directed rotameric approach. *Proc. Natl Acad. Sci. USA* **107**, 8177–8182 (2010).
76. Chou, F. C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10**, 74–76 (2013).
77. Afonine, P. V., Headd, J. J., Terwilliger, T. C. & Adams, P. D. New tool: phenix.real_space_refine. *Computational Crystallography Newsletter* 43–44 (2013).
78. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
79. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
80. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
81. Bock, L. V. *et al.* Energy barriers and driving forces in tRNA translocation through the ribosome. *Nat. Struct. Mol. Biol.* **20**, 1390–1396 (2013).
82. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
83. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
84. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
85. Joung, I. S. & Cheatham, T. E., III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).
86. Aduri, R. *et al.* AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. *J. Chem. Theory Comput.* **3**, 1464–1475 (2007).
87. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260 (2006).
88. Amadei, A., Linssen, A. B. & Berendsen, H. J. Essential dynamics of proteins. *Proteins* **17**, 412–425 (1993).
89. Schrader, J. M. & Uhlenbeck, O. C. Is the sequence-specific binding of aminoacyl-tRNAs by EF-Tu universal among bacteria? *Nucleic Acids Res.* **39**, 9746–9758 (2011).
90. Haruna, K., Alkazemi, M. H., Liu, Y., Söll, D. & Englert, M. Engineering the elongation factor Tu for efficient selenoprotein synthesis. *Nucleic Acids Res.* **42**, 9976–9983 (2014).
91. Itoh, Y. *et al.* Decameric SelA•tRNA(Sec) ring structure reveals mechanism of bacterial selenocysteine formation. *Science* **340**, 75–78 (2013).
92. Li, W. *et al.* Recognition of aminoacyl-tRNA: a common molecular mechanism revealed by cryo-EM. *EMBO J.* **27**, 3322–3331 (2008).
93. Takyar, S., Hickerson, R. P. & Noller, H. F. mRNA helicase activity of the ribosome. *Cell* **120**, 49–58 (2005).
94. Budkevich, T. V. *et al.* Regulation of the mammalian elongation cycle by subunit rolling: a eukaryotic-specific ribosome rearrangement. *Cell* **158**, 121–131 (2014).
95. Frank, J. & Agrawal, R. K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406**, 318–322 (2000).



Extended Data Figure 1 | Computational sorting of ribosome particle images. **a**, Hierarchical sorting scheme. Numbers refer to classification steps. SVC, supervised classification by projection matching; "bad" 70S, the particles of low quality and/or those showing mixtures of different compositions and conformations; initial complex (hybrid) (IC (hybrid)), ratcheted state of the ribosome with one tRNA in the hybrid state. **b**, Mask

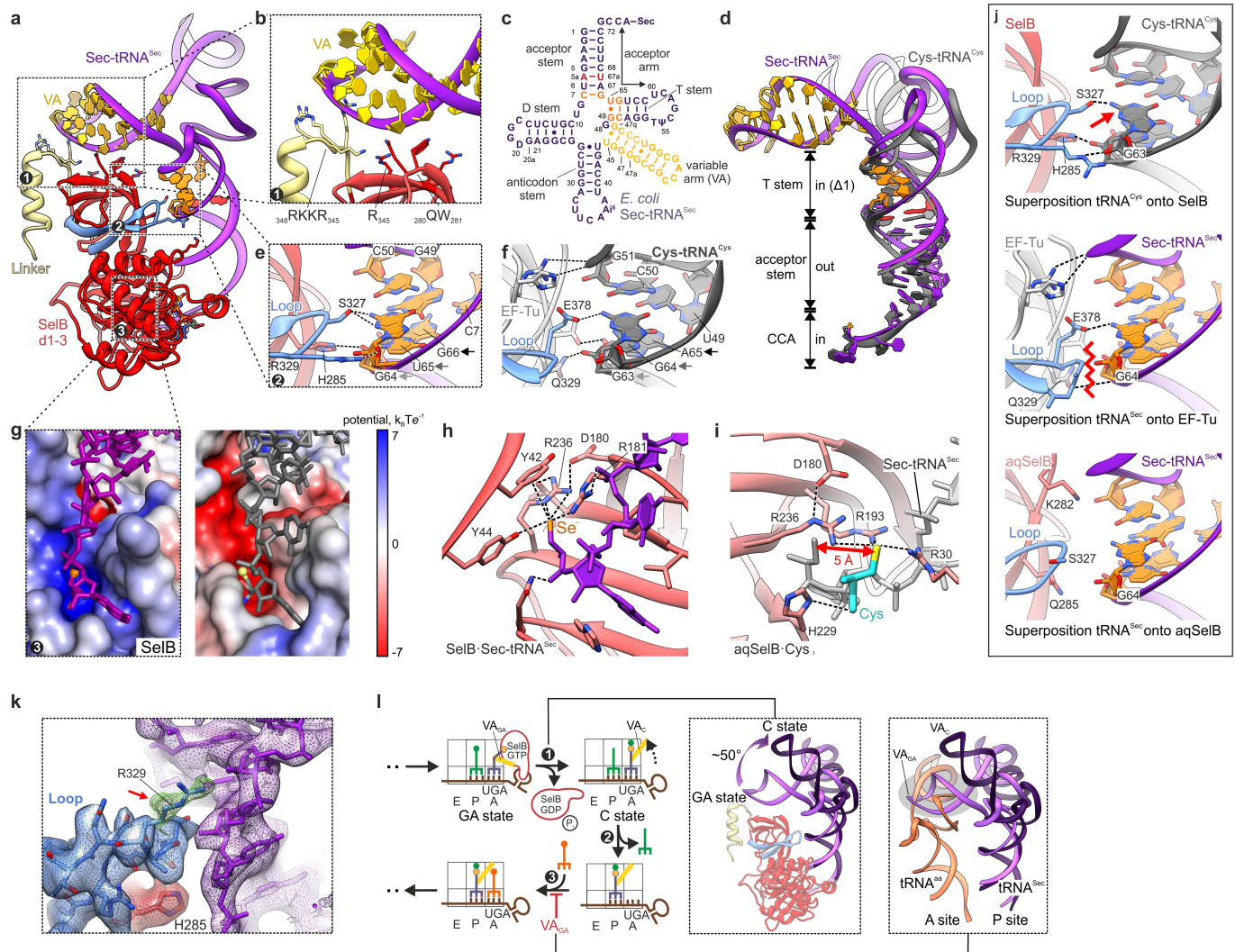
used for focused classification in step 4. **c**, Cryo-EM reconstruction from 100,000 random unsorted particle images. Left, initial refinement stage at 10 Å resolution showing only scattered density (red) for SelB, Sec-tRNA^{Sec} and SECIS. Right, final cryo-EM map at 3.7 Å resolution. **d**, Structural changes resolved by computational sorting as quantified by the r.m.s.d. and the changes in mass due to ligand binding and dissociation.



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Cryo-EM map and model quality. **a**, Fourier shell correlation (FSC) curves and model validation. FSC curves are shown for each state. Black, the FSC curve computed between the masked independent half-maps from cryo-EM refinement (half1 and half2); blue, the FSC curve between the final cryo-EM map (map) and the final model (model); red, the FSC curve between half-map 1 (half1) and the model obtained by refinement only against half-map 2 (model2). The vertical black dashed line indicates the resolution according to the 0.143 criterion (grey line) and the maximum resolution at which the full atomic models were refined. **b**, FSC curves (FSC_{work}) computed for each state between reflections from solvent-flattened cryo-EM maps and model maps generated from refined atomic coordinates. Differences to **d** largely result from solvent-flattening (Methods). **c**, CC_{work} curves as obtained by refinement for each state. For reliable resolution estimates CC_{work} is expected to be >0.2 in the highest resolution shell. **d**, Cryo-EM densities and models of the quarternary complex, SelB–GDPNP–Sec-tRNA^{Sec}–SECIS-mRNA for each SelB-bound ribosome state. Densities are coloured

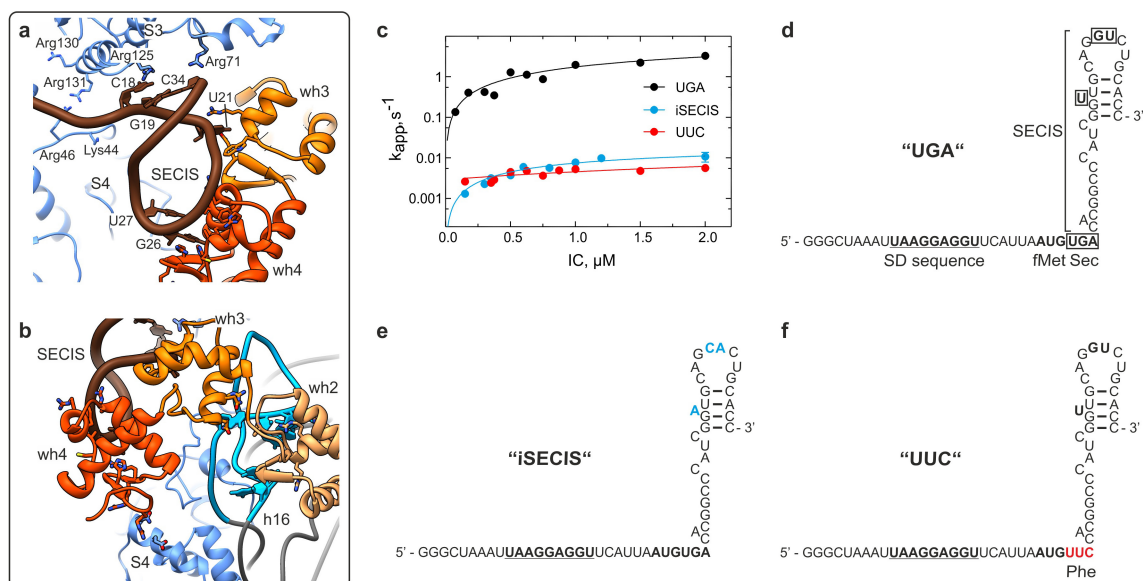
as in Fig. 1b and rendered at $\sim 2\sigma$, except wh3, wh4 and SECIS of the GTPase-activated state, which are rendered at $\sim 1.5\sigma$. Top right: Sec-tRNA^{Sec} density (purple) in the initial binding state at 5.3 Å resolution; the conformation of the invariant histidine 61 is not discernible (question mark). Middle right: Density for histidine 61 (blue) in the codon reading state indicating a partially flipped-in conformation. **e**, Densities for rRNA modifications as seen in the GTPase-activated state at 3.4 Å resolution. Densities are rendered at $\sim 3\sigma$. Arrows denote the characteristic distortion of the nucleobase of D2449 and methyl groups for the other modified nucleotides. **f**, **g**, Cryo-EM reconstructions of the individual states (as indicated) coloured according to local resolution. Left: Surface view; right: Cut-away view. Heat maps are adjusted to the respective resolutions ranges of the cryo-EM maps in **f** and **g**. The arrow denotes the substantially lower local resolution of SelB residues 485 to 614 in the GTPase-activated state; atomic models for such regions were correspondingly refined at lower resolution (Methods).



Extended Data Figure 3 | See next page for caption.

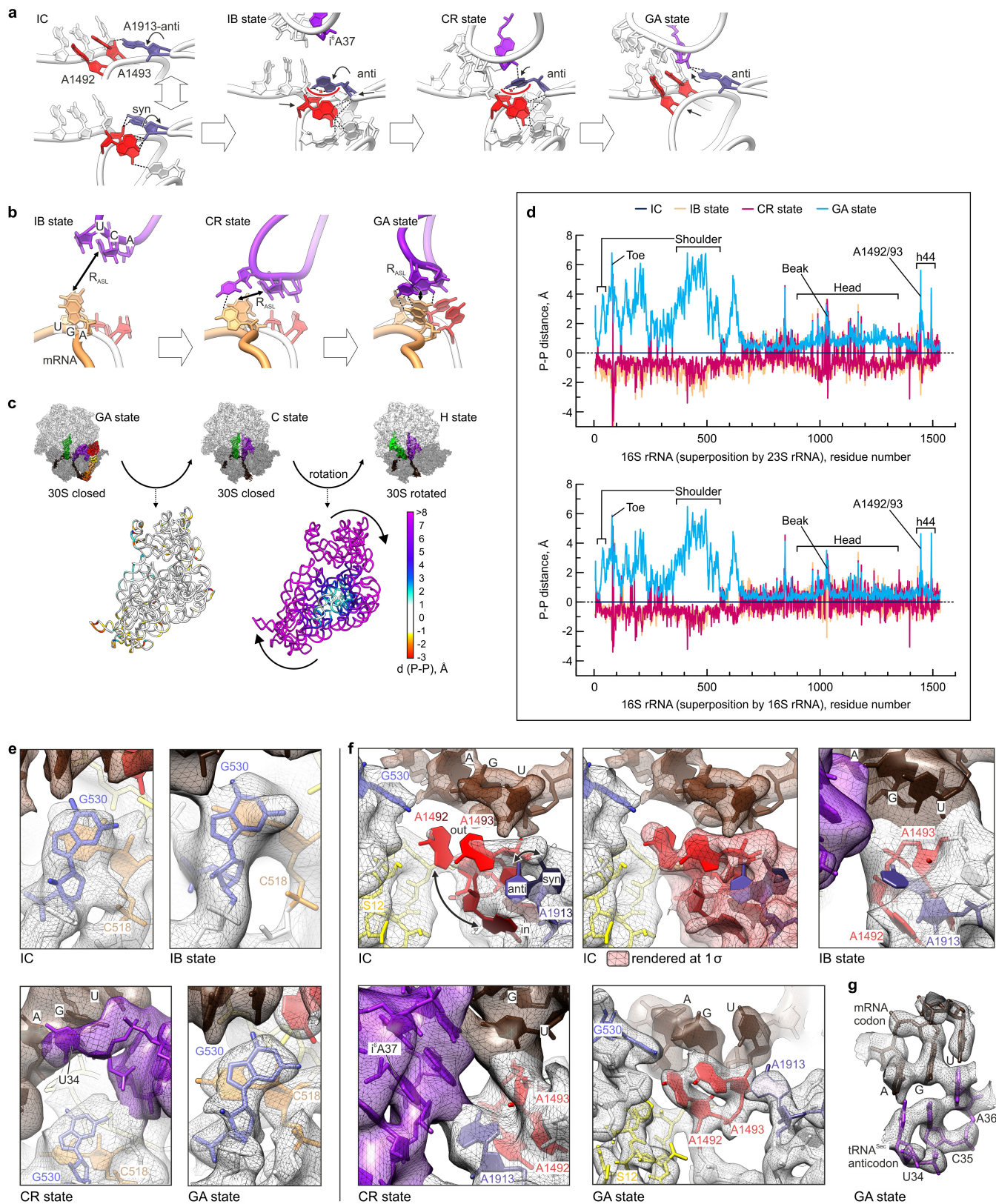
Extended Data Figure 3 | Structural basis for the specificity of tRNA^{Sec} recognition by SelB. All superpositions are based on domains 1 and 2 of SelB and EF-Tu, respectively. **a**, Overview of SelB-GDPNP-Sec-tRNA^{Sec} bound to the ribosome in the GTPase-activated state. Numbers refer to the close-ups in **b**, **e** and **g**, respectively, which depict specific SelB-Sec-tRNA^{Sec} interactions. SelB d1–3, domains 1 to 3 of SelB; VA, variable arm. **b**, Interactions between SelB (domain 3, red, and the linker, residues 340–363 in khaki) and the backbone of the extra-long variable arm of tRNA^{Sec}. These interactions do not form in the canonical EF-Tu-aa-tRNA complexes. **c**, The structure of tRNA^{Sec}. In contrast to canonical tRNAs, tRNA^{Sec} entails an additional base-pair in the acceptor stem (A5a–U67a, red), a unique sequence in the T-stem, the so-called antideterminant box (C7–G66, G49–U65 and C50–G64, orange), and a very long variable arm (yellow). **d**, Superposition of tRNA^{Sec} (purple) and the canonical tRNA^{Cys} (dark grey) in the complex with SelB-GDPNP and EF-Tu-GDPNP, respectively. The structure of the canonical EF-Tu-GDPNP-Cys-tRNA^{Cys} is from PDB structure 1B23 (ref. 11). Note the distortion in the acceptor stem of tRNA^{Sec} (residues G1–U6/G72–A67) as compared to tRNA^{Cys} that compensates for the additional base pair. As a result of the distortion, the main factor-binding sites of both tRNAs, the CCA end and the T-stem, overlap and the bases are arranged in register ('in'). In the T-stem, the bases are shifted by exactly one base pair ('Δ1') between the two tRNAs, which shifts the bases out of register ('out') in the acceptor stems of the two tRNAs. **e**, Specific interaction between the extended loop (blue) in SelB domain 3 and the antideterminant box in the T-stem of tRNA^{Sec}. **f**, Sequence-specific interaction between the loop (blue) in EF-Tu domain 3 and the T-stem of tRNA^{Cys} (PDB ID 1B23 ref. 11). Note the different structure of the hairpin-loop as compared to SelB, contributing to the particularly low affinity of EF-Tu for tRNA^{Sec} (ref. 89) (see also **k**, middle panel). **g**, Charge distribution in the amino-acid-binding pockets of SelB (left) and EF-Tu (right). The colour bar on the right denotes the electrostatic potential in $k_B T e^{-1}$. Note the similar geometry of tRNA binding in the two complexes. In SelB, the pocket is positively charged, facilitating the interaction with the negatively charged selenol group of Sec-tRNA^{Sec} and the discrimination against the precursors of Sec-tRNA^{Sec} biosynthesis, deacylated tRNA^{Sec} and Ser-tRNA^{Sec}, as well as against most canonical tRNAs. In EF-Tu, the amino-acid-binding pocket is negatively charged, thereby discriminating against Sec-tRNA^{Sec}. **h**, Recognition of the selenol group in Sec-tRNA^{Sec} by SelB. The positively charged, highly conserved Arg181 and Arg236 in SelB contact the

negatively charged selenol group (Se⁻, orange), whereas the aromatic ring of Tyr42 stacks onto the selenol group. The importance of these residues was demonstrated by mutational analysis^{9,90}. The universally conserved Asp180, which is also important for Sec-tRNA^{Sec} binding⁹, forms a secondary binding shell stabilizing Arg236. **i**, Solvent-exposed amino-acid-binding site as found in the crystal structure of isolated *Aquifex aeolicus* (aq) SelB-GDPNP co-crystallized with L-cysteine as Sec mimic (PDB ID 4ZU9)⁸. The position of Cys deviates by 5 Å from the amino acid position in the present SelB-GDPNP-Sec-tRNA^{Sec} structure (grey), resulting in a distinct interaction pattern and a more solvent-exposed binding site of the Cys. **j**, Superpositions of SelB-GDPNP-Sec-tRNA^{Sec} and EF-Tu-GDPNP-Cys-tRNA^{Cys} revealing differences in T-stem recognition by the loop in SelB domain 3 versus EF-Tu. Top: The loop structure in SelB is compatible with binding of canonical tRNA^{Cys}. Other tRNAs with different T stem sequences (particularly at position 63, red arrow) may be unfavourable for the interaction with the conserved Ser327 in SelB. Middle: The loop structure in EF-Tu is incompatible with the conformation of the antideterminant box (orange) in tRNA^{Sec}. Bottom: In isolated aqSelB (PDB ID 4ZU9)⁸, the loop adopts a similar conformation as in the SelB ternary complex, suggesting a conserved loop structure that facilitates recognition of tRNA^{Sec}. **k**, Experimental density depicting the interaction of SelB with the acceptor and T stem region of tRNA^{Sec} in the GTPase-activated state (rendered at 2σ) suggesting alternative conformations of Arg329; green mesh, density rendered at 1σ. **l**, Scheme depicting the functional importance of rearrangements of the tRNA^{Sec} variable arm for translation elongation after Sec incorporation. (1) Upon release from SelB, tRNA^{Sec} accommodates in the aminoacyl site in the classical state. The variable arm undergoes a large reorientation from a SelB-bound distorted conformation in the GTPase-activated state to its free ground state conformation in the classical state (left close-up); the latter ground state conformation is similar to the conformation of the variable arm found in complexes of tRNA^{Sec} with its conversion enzymes, seryl-tRNA synthetase⁶ and Sela⁹¹. (2) After peptide bond formation (resulting in the hybrid state, not shown), peptidyl-tRNA^{Sec} is translocated to the classical peptidyl site. (3) Binding of the next aa-tRNA (orange) is only compatible with the variable arm adopting the classical state conformation; the GTPase-activated state-conformation of the variable arm would sterically interfere with binding of the next aa-tRNA (right close-up). The long variable arm of canonical class 2 aa-tRNA is not distorted upon EF-Tu binding⁹² and, consequently, does not require any changes during translation elongation.



Extended Data Figure 4 | The SelB-SECIS interactions on the ribosome and kinetic analysis of GTP hydrolysis. **a**, Close-up of the SECIS-ribosome interactions in the GTPase-activated state. Ribosomal proteins S3, S4, and S5 contribute to the helicase activity of the ribosome⁹³. Arg130 and Arg131 of S3 and Lys44 and Arg46 of S4 are important for mRNA unwinding⁹³. Other potential helicase elements are Arg71 and Arg125 in S3, which interact with the base of the SECIS hairpin and may thereby facilitate unwinding of the mRNA secondary structure elements. The mode of SECIS recognition by SelB on the ribosome is similar to that in isolated SelB domain 4-SECIS complexes, including three essential bulged nucleotides in the SECIS (U21, G26 and U27; present data and refs 9, 25 and 72). **b**, Interaction of SelB domain 4 with the shoulder of

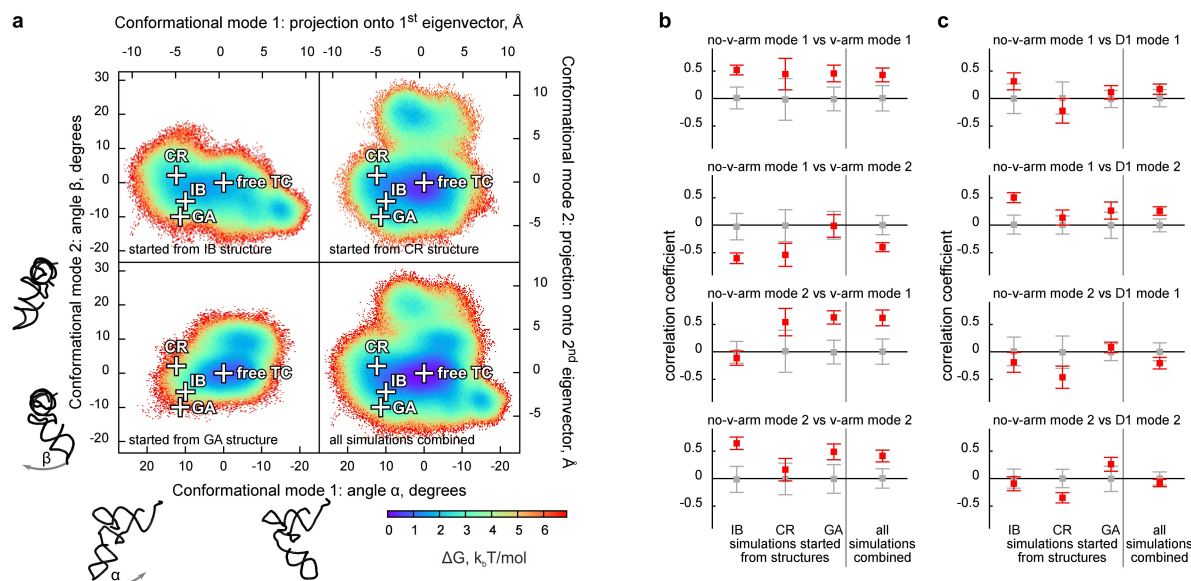
the 30S subunit. Winged-helix motif 4 (wh4) of SelB forms salt bridges with protein S4, while wh2 and wh3 embrace helix 16 of 16S rRNA by interacting mainly with the rRNA backbone. **c**, Concentration dependence of the apparent rate of GTP hydrolysis by SelB measured with increasing concentrations of ribosomes programmed with mRNAs 'UGA' (black), 'iSECIS' (blue) and 'UUC' (red). **d**, mRNA construct 'UGA' used for structural and kinetic analyses containing the cognate UGA codon coding for Sec and the functional minimal *fdhF*-SECIS. The Sec codon and the essential bases of the SECIS are indicated by boxes; SD, Shine-Dalgarno sequence. **e**, **f**, mRNA constructs 'iSECIS' and 'UUC' used for kinetic analysis. Changes in comparison to the 'UGA' mRNA are indicated in blue (iSECIS) and red (UUC).



Extended Data Figure 5 | See next page for caption.

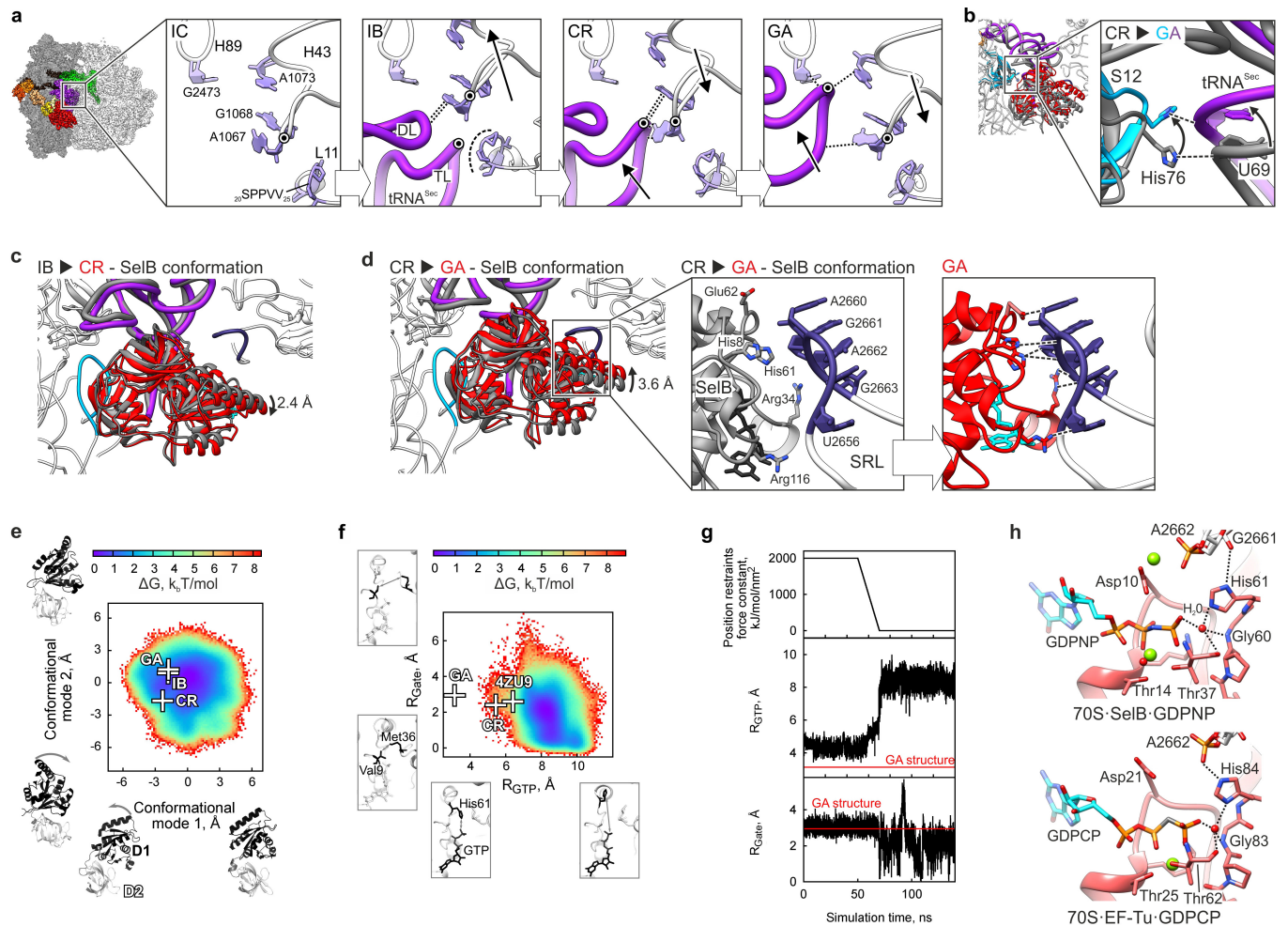
Extended Data Figure 5 | Local and global conformational changes of the 30S subunit. **a**, Rearrangements of A1913 in H69 of 23S rRNA. A1913 stabilizes A1492 and A1493 and guides Sec-tRNA^{Sec} to the mRNA codon by interaction with isopentenyl-A37 (i6A37) of tRNA^{Sec}. The respective conformation (syn or anti) is denoted; dashed lines indicate potential interactions with distances ≤ 4 Å. **b**, Close-up of the distance R_{ASL} between the UGA (N1 of G) and anticodon of tRNA^{Sec} (N3 of C35). Note the potential interaction between the third codon position and the tRNA anticodon in the codon reading state. **c**, Global conformational changes of the 30S subunit upon accommodation (GTPase-activated to classical) and hybrid state formation (classical to hybrid). tRNA accommodation in the aminoacyl site does not cause major 30S subunit rearrangements, in contrast to the eukaryotic system, where tRNA accommodation correlates with a large-scale conformational change⁹⁴. Subsequent tRNA hybrid state formation is coupled to the well-known rotational movement of the 30S subunit⁹⁵. The typical rotational changes all over the 30S subunit are clearly distinct from the changes observed upon domain opening and closure, which are mainly found at the shoulder region (Fig. 3a). The heat map quantifies the movements of 16S rRNA backbone phosphate atoms as obtained by superposition on 23S rRNA of the initial complex state (Fig. 3a). **d**, Deviations of the 16S rRNA backbone phosphates in the initial binding, codon reading, or GTPase-activated states from the initial complex superimposed on 23S rRNA (top) or 16S rRNA (bottom).

Negative values correspond to the 30S domain opening, positive values to domain closure. Landmarks of the 30S subunit are indicated. **e**, **f**, Experimental densities of the decoding centre region rendered at 2–2.5 σ , if not indicated otherwise. For better visibility, densities for tRNA^{Sec} were omitted for the GTPase-activated state. **e**, Anti-conformation of G530 and stacking interaction with C518 of 16S rRNA as seen in the different states. Bottom left: Note the density of tRNA^{Sec} (purple) suggesting a partial interaction of U34 with the mRNA codon in the codon reading state. **f**, Different conformational states of A1492, A1493 (red) of 16S rRNA and A1913 (slate blue) of 23S rRNA. Top left: In the initial complex the reduced densities for the three nucleotides indicate a dynamic equilibrium between two states: (1) A1492, A1493 flipped out ('out', red) with A1913 in anti-configuration ('anti', slate blue) and (2) A1492, A1493 flipped in ('in', dark red) with A1913 in syn-configuration ('syn', dark slate blue). Top middle: The dynamic nucleotides in the initial complex are discernible at lower threshold (red mesh, density at $\sim 1\sigma$). Top right and bottom left: Density in the initial binding and codon reading states is compatible with the bases of A1492 and A1493 flipped-in and A1913 in the anti-configuration. Bottom right: In the GTPase-activated state, A1492 and A1493 are flipped-out and A1913 remains in the anti-conformation. **g**, Density for the codon–anticodon interaction in the GTPase-activated state rendered at $\sim 3\sigma$.



Extended Data Figure 6 | Molecular dynamics simulations of the free ternary complex. **a**, tRNA^{Sec} dynamics obtained by molecular dynamics simulations of the free SelB–GTP–Sec-tRNA^{Sec} ternary complex. The free energy landscape for the conformational dynamics of the tRNA^{Sec} body, excluding the variable arm, is shown as a heat map. The free energy was estimated from separate simulations starting from the structures of the ternary complex bound to the ribosome in initial binding, codon reading, and GTPase-activated states (upper and lower left panels), as well as from all simulations combined (lower right panel). The results of different simulations have a large overlap, especially in regions of low free energy. The crosses denote the tRNA conformations in the ribosome-bound cryo-EM states and the free energy minimum of the free ternary complex found in the molecular dynamics simulations. **b**, Weak coupling between the conformational dynamics of tRNA^{Sec} excluding the variable

arm (no-v-arm) and of the variable arm (v-arm). Each panel shows the correlation coefficients for a pair of conformational modes (red points, with standard deviation) for the simulations starting from each state (initial binding, codon reading, and GTPase-activated), as well as for all simulations combined. The grey bars denote the standard deviation of the correlation coefficient expected from random drawing of projections. All pairs of modes show a small but significant correlation, indicating a weak coupling. **c**, Coupling between tRNA dynamics (no-v-arm modes) and SelB motions (SelB domain 1, D1, relative to domain 2, D2). Only a weak overall correlation is seen between no-v-arm mode 1 and D1 mode 2. The three ribosome-bound conformations do not change much with regards to no-v-arm mode 1 (Fig. 3b), which suggests that there is no direct coupling between tRNA dynamics and SelB motions on the ribosome.



Extended Data Figure 7 | Role of L11-rRNA arm, protein S12 and the dynamics of SelB domains. **a**, Role of L11-rRNA arm in guiding Sec-tRNA^{Sec} towards H89. Residues of protein L11, H43 and H89 of 23S rRNA are shown in mauve; dashed lines mark interactions with the D loop (DL) and T loop (TL) of tRNA^{Sec} (purple); arrows denote changes with respect to the preceding state. **b**, Protein S12 as a pivot for tRNA^{Sec} movement. S12 His76 interacts with the backbone of tRNA^{Sec} (residue U69) guiding the tRNA from its position in the codon reading state (grey) to the GTPase-activated state (coloured). **c**, Inter-subunit rearrangement in SelB upon transition from initial binding to codon reading. Shown is the small movement observed in addition to the rotation of SelB relative to the 30S shoulder as depicted in Fig. 4f, upper panel. **d**, SelB residues facilitating SRL docking. Left and middle: Spontaneous fluctuations within SelB that are required—in addition to the rotation upon 30S domain closure (shown in Fig. 4, lower panel)—for transition from codon reading (grey) to the GTPase-activated state (red). Middle: 30S domain closure and tRNA^{Sec} repositioning re-align SelB and in particular its long polar residues towards the SRL. Right: SelB-SRL interactions resulting from docking in the GTPase-activated state. **e**, Motions of SelB domain 1 (D1) relative to domain 2 (D2) in the molecular dynamics simulations of the free

ternary complex. SelB domain 1 can rapidly and spontaneously sample arrangements required for the transitions from initial binding to codon reading and GTPase-activated states. **f**, Minimum distance between the imidazole ring of SelB His61 and γ -phosphate/oxygens of GTP (R_{GTP}) and the minimum distance between the Van der Waals radii of Val9 and Met36 (R_{Gate}), residues adjacent to His61 in SelB and analogous to the presumed 'gate' in EF-Tu³. In the free ternary complex, the distance between GTP and His61 of SelB is larger than in the GTPase-activated state on the ribosome, although Val9 and Met36 can freely move apart. Distances for the crystal structure of SelB-GDPNP are labelled 4ZU9 (ref. 8). **g**, Metastable SelB conformation in the GTPase-activated state. SelB in the GTPase-activated state was taken as initial structure for molecular dynamics simulations. During the first 50 ns of the simulations, the positions of heavy atoms are restrained to allow equilibration of the solvent (top panel). As soon as the restraints are released (50–70 ns), the distance between His61 and GTP increases (middle panel), and residues Val9 and Met36 resume fluctuations between open and closed conformations. **h**, Comparison of the active sites in SelB and EF-Tu. Top: GTPase centre of SelB in the GTPase-activated state. Bottom: Activated GTPase centre of EF-Tu on the ribosome (PDB ID 4V5L)³⁵.

Extended Data Table 1 | Data collection and model refinement

Ribosomal state	IC	IB state	CR state	GA state	C state	H state
Database entries						
EMDB ID	4121	4122	4123	4124	4125	4126
PDB ID	5LZA	5LZB	5LZC	5LZD	5LZE	5LZF
Data collection						
Final resolution (Å)	3.6	5.3	4.8	3.4	3.5	4.6
Particles	73,598	8,002	11,658	159,729	130,705	19,868
Pixel size (Å)	1.16	1.16	1.16	1.16	1.16	1.16
Defocus range (μm)	0.7-2.6	0.7-2.6	0.7-2.6	0.7-2.6	0.7-2.6	0.7-2.6
Electron dose (e [−] /Å ²)	30±5	30±5	30±5	30±5	30±5	30±5
Reciprocal space data						
Space group	P1	P1	P1	P1	P1	P1
a, b, c (Å) *	315.6, 315.6, 315.6	334.1, 334.1, 334.1	334.1, 334.1, 334.1	315.6, 315.6, 315.6	315.6, 315.6, 315.6	334.1, 334.1, 334.1
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wilson B (Å ²)	83.75	158.2	145.4	69.8	76.6	165.1
Refinement[†]						
Resolution range (Å) ‡	223.16 – 3.60 (3.64 – 3.60)	236.23 – 5.30 (5.36 – 5.30)	236.23 – 4.80 (4.85 – 4.80)	223.16 – 3.40 (3.44 – 3.40)	223.16 – 3.50 (3.54 – 3.50)	223.23 – 4.60 (4.65 – 4.60)
No. Reflections	1409938 (112206)	523391 (45504)	705417 (23029)	1674406 (54565)	1534972 (50138)	801725(26303)
R _{work} /CC _{work} [§]	0.270/0.929 (0.470/0.342)	0.306/0.954 (0.500/0.255)	0.317/0.947 (0.496/0.250)	0.255/0.935 (0.468/0.411)	0.252/0.935 (0.448/0.430)	0.290/0.956 (0.486/0.305)
FSC _{work} [¶]	0.924 (0.608)	0.927 (0.516)	0.919 (0.509)	0.934 (0.700)	0.933 (0.668)	0.936 (0.573)
Cumulative RSCC* (%)	68/92/95	68/93/97	55/90/96	73/92/96	72/92/96	72/93/97
>0.8/>0.6/>0.4						
ML based phase error (°)	31.33	41.6	42.3	28.7	28.2	36.9
Molprobit all-atom clashscore	9.9	7.0	6.8	7.0	10.34	5.6
Residues with atoms out of density for main chain/side chain*	0/1	9/0	7/0	0/0	0/1	3/0
No. Atoms/RSCC ^{¶¶}	147012/0.805	153989/0.826	153989/0.793	154136/0.834	149015/0.814	149002/0.828
Protein	46375/0.776	51254/0.782	51254/0.760	51254/0.809	46375/0.786	46424/0.802
RNA, tRNA	100635/0.813	102698/0.834	102698/0.801	102698/0.841	102638/0.822	102576/0.835
Ligand, ion	2	37	37	184	2	2
B-factors (Å ²)	107.6	247.0	189.1	106.4	103.7	198.0
Protein	127.1	274.0	212.9	127.1	121.6	221.0
RNA, tRNA	98.7	233.5	177.2	96.1	95.6	187.6
Ligand, ion	100.5	385.8	289.3	60.0	104.8	268.0
R.m.s deviations						
Bond lengths (Å)	0.012	0.011	0.011	0.011	0.012	0.009
Bond angles (°)	1.220	1.350	1.285	1.301	1.238	1.236
Ramachandran plot statistics						
Most favored (%)	90.29	92.12	92.15	90.83	90.05	90.81
Disallowed (%)	1.56	1.93	1.83	1.30	1.97	1.93
Avg. RSCC* for tRNA ^{Sec} /SECIS	n.a./0.12**††	0.72/0.70	0.58/0.57	0.72/0.41††	0.61/0.20††	0.63/0.35††
Avg. RSCC* for SelB d1-3/wh1-2/wh3-4	n.a.	0.68/0.73/0.66	0.58/0.70/0.52	0.81/0.78/0.39††	n.a.	n.a.

*For refinement, maps at ≥4.6 Å resolution were cropped to 280 × 280 × 280 pixels; maps at ≤3.6 Å resolution were resampled to 400 × 400 × 400 pixels, corresponding to a pixel size of 0.789 Å.

†Refinement target: MLHL maximum likelihood with experimental phase probability distribution.

‡Highest resolution shell is shown in parenthesis.

§ $R_{work} = \sum |F_{EM}| - |F_{MODEL}| / \sum |F_{EM}|$, where F_{EM} are structure factors calculated on the basis of solvent flattened EM map and F_{MODEL} are structure factors calculated from the refined model. The structure factors belonged to the working set which was used for reciprocal space refinement.

¶ CC_{work} = Pearson correlation coefficient calculated between F_{EM} and F_{MODEL} .

¶¶FSC_{work} is averaged over all FSC shells (FSC_{overall}) calculated between F_{EM} and F_{MODEL} belonging to the working set. The value for the highest resolution shell is shown in parenthesis and was calculated using the form: $FSC_{work(shell)} = \sum (F_{EM} \times F_{MODEL} \cos(\Delta phase)) / ((\sum |F_{EM}|^2) \times (\sum |F_{MODEL}|^2))$.

#Residue-averaged real space local correlation coefficient in region of model (RSCC) to EM map calculated with RESOLVE⁷⁸.

*Calculated with RESOLVE⁷⁸.

**For the initial complex, the average RSCC for the SECIS was computed against the map filtered to lower resolution; the resulting low RSCC value indicates substantial conformation variability of the SECIS.

††Modelled based on cryo-EM map filtered to lower resolution.

Observed glacier and volatile distribution on Pluto from atmosphere–topography processes

Tanguy Bertrand¹ & François Forget¹

Pluto has a variety of surface frosts and landforms as well as a complex atmosphere¹. There is ongoing geological activity related to the massive Sputnik Planitia glacier, mostly made of nitrogen (N₂) ice mixed with solid carbon monoxide and methane², covering the 4-kilometre-deep, 1,000-kilometre-wide basin of Sputnik Planitia^{1,3} near the anti-Charon point. The glacier has been suggested to arise from a source region connected to the deep interior, or from a sink collecting the volatiles released planetwide¹. Thin deposits of N₂ frost, however, were also detected at mid-northern latitudes and methane ice was observed to cover most of Pluto except for the darker, frost-free equatorial regions². Here we report numerical simulations of the evolution of N₂, methane and carbon monoxide on Pluto over thousands of years. The model predicts N₂ ice accumulation in the deepest low-latitude basin and the threefold increase in atmospheric pressure that has been observed to occur since 1988^{4–6}. This points to atmospheric–topographic processes as the origin of Sputnik Planitia's N₂ glacier. The same simulations also reproduce the observed quantities of volatiles in the atmosphere and show frosts of methane, and sometimes N₂, that seasonally cover the mid- and high latitudes, explaining the bright northern polar cap reported in the 1990s^{7,8} and the observed ice distribution in 2015². The model also predicts that most of these seasonal frosts should disappear in the next decade.

To understand the distribution of volatiles on Pluto, we developed a numerical volatile transport model designed to represent the physical processes that control their condensation, sublimation and atmospheric transport forced by the variation of insolation throughout multiple annual cycles (see Methods). The model derives from a full Pluto General Circulation Model⁹ (GCM) able to calculate the three-dimensional transport and mixing in the atmosphere, but too slow to run for multiple Pluto seasons. To simulate the cycles of the volatiles for thousands of terrestrial years in a practical amount of computing time, atmospheric dynamics and transport were parameterized using a simplified redistribution of N₂, carbon monoxide (CO) and methane (CH₄) gases with characteristic timescales estimated from 'short' GCM simulations. We start our simulations with Pluto uniformly covered with 50 kg m^{−2} of each kind of ice, and let the modelled planet evolve for 50,000 terrestrial years. The seasonal ice cycles are then repeatable from one Pluto year to the next. However, they strongly depend on key model parameters such as the topography, the albedo and the emissivity of the ices (only partly constrained by observation and theory), the total ice inventories, and the thermal conductivity of the shallow (centimetres) and deep (metres) subsurface, which control the diurnal¹⁰ and seasonal global thermal inertia, respectively.

Assuming a flat surface, the model reproduces the results of pre-New Horizons N₂ cycle models^{11,12}. Like those models, we find that

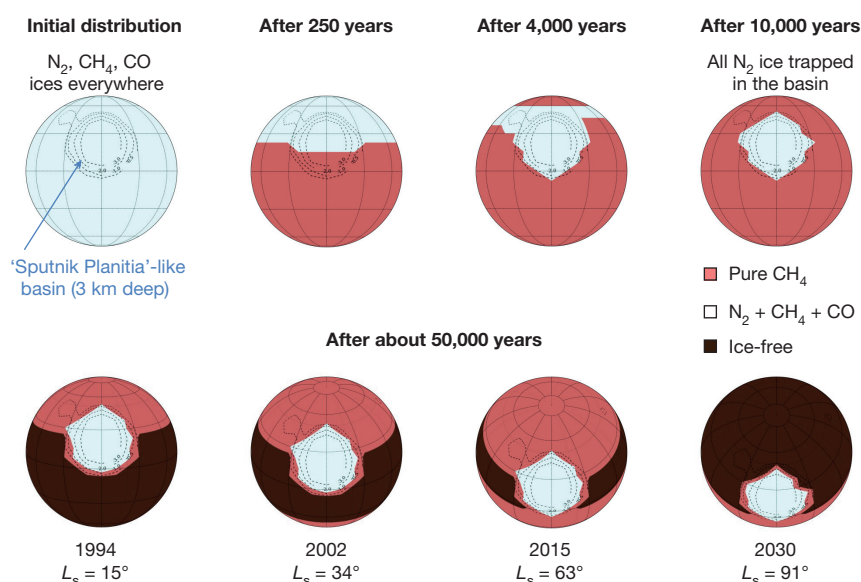


Figure 1 | Surface maps from the reference simulation with $TI = 800 \text{ J s}^{-1/2} \text{ m}^{-2} \text{ K}^{-1}$. The top row shows the initial 10,000 years, starting with all ices uniformly distributed. N₂ ice is sequestered in the basin after 10,000 Earth years. The bottom row shows results after 50,000 Earth years, with methane ice in a quasi-equilibrated state. The maps are shown for 1994, 2002 and 2015 for the sake of comparison with Hubble

Telescope^{7,8,18} and New Horizons^{1,2} data. The corresponding solar longitudes L_s (that is, the Pluto–Sun angle measured from the Northern spring equinox where $L_s = 0^\circ$) are shown. The sub-latitude shown is the sub-solar point and the central longitude is the anti-Charon point (Sputnik Planitia).

¹Laboratoire de Météorologie Dynamique, IPSL, Sorbonne Universités, UPMC Université Paris 06, CNRS, BP99, 4 place Jussieu, 75005 Paris, France.

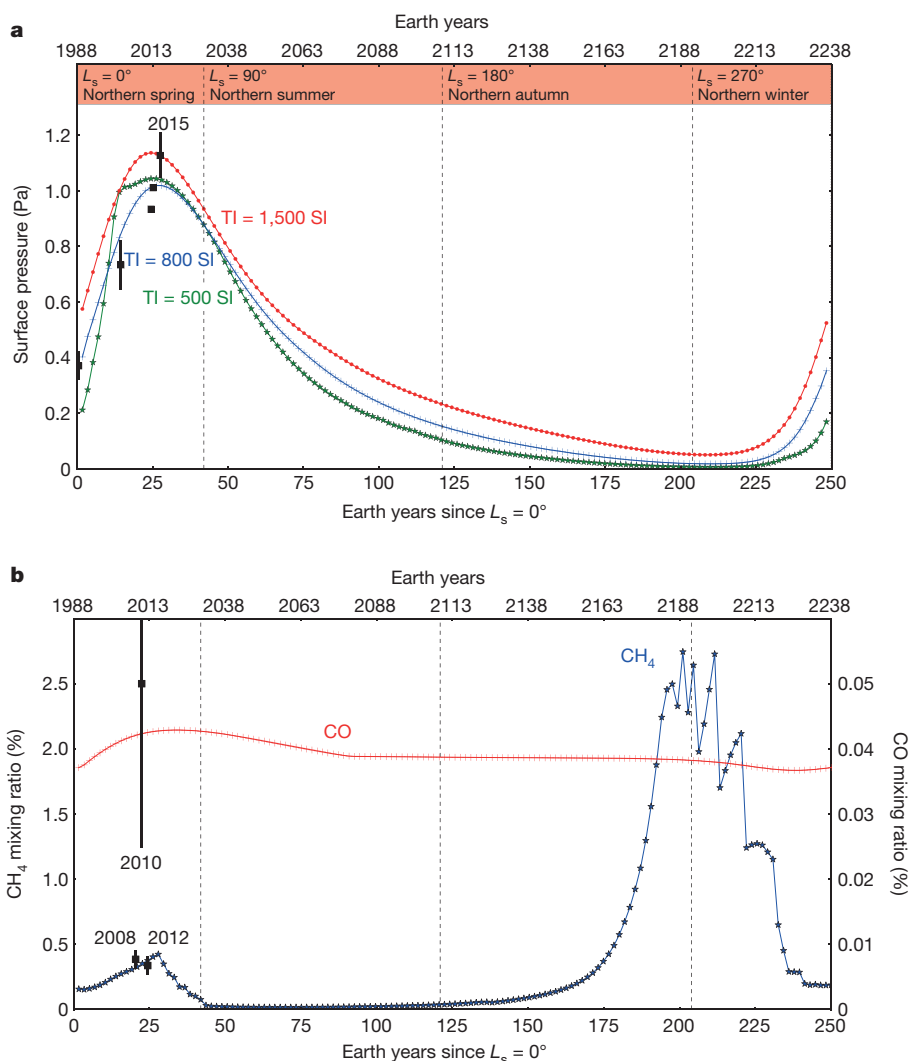


Figure 2 | Modelled annual evolution of atmospheric volatiles.

a, Surface pressure P_s for three TI scenarios. For each value of TI, the N_2 ice albedo and emissivity were chosen to yield a surface pressure near 1–1.1 Pa in July 2015 (see Methods). Black squares show P_s measured by stellar occultations⁶ at 1,215 km (1σ) from Pluto's centre ($P_{1,215\text{ km}}$),

assuming the same ratio $P_s/P_{1,215\text{ km}}$ as measured in 2015. **b**, Global mean mixing ratio of atmospheric CO and CH₄ obtained with the reference simulation (TI = 800 J s^{-1/2} m⁻² K⁻¹). The 2008–2012 values are consistent with terrestrial observations^{17,19} (black squares; see Methods).

seasonal thermal inertia (TI, measured in units of J s^{-1/2} m⁻² K⁻¹), a poorly known parameter, is the key driver of the N_2 cycle. On a flat Pluto, two end-member behaviours arise. High TI leads to the formation of a permanent band of N_2 ice in the equatorial regions because the annual-mean insolation is lower there than at the poles, as a result of the high obliquity of Pluto¹³. On the other hand, low TI yields more pronounced seasonal variations and the formation of N_2 seasonal polar caps only.

We examined the effect of topography, whose origin on Pluto we do not discuss here, by placing a 4-km-deep circular crater at the location of Sputnik Planitia and two smaller craters corresponding to the informally called Burney (1,000 m deep) and Guest (800 m deep) craters¹.

With such an orography, in the high-TI cases, all the N_2 ice is entirely sequestered in the modelled 'Sputnik Planitia'-like basin after roughly 10,000 Earth years (Fig. 1). The N_2 ice does not form a permanent latitudinal band, as in the flat planet case, because the basin induces a higher surface pressure and thus a higher condensation temperature, which leads to a stronger thermal infrared cooling and a higher condensation rate. This phenomenon is also observed on Mars, where CO₂ frosts preferably form at low elevations such as the Hellas basin¹⁴ (Extended Data Fig. 1). The characteristics of Sputnik Planitia are thus

explained by its latitude and depth rather than by a connection with putative N_2 reservoirs in the deep interior.

In our model, N_2 ice is entirely trapped in Sputnik Planitia when using a TI value higher than 700 J s^{-1/2} m⁻² K⁻¹ coupled with an ice albedo up to 0.7 and an emissivity up to 0.8. Decreasing TI to fall in the range 400–700 J s^{-1/2} m⁻² K⁻¹ allows surface temperatures to reach the N_2 condensation point at other locations during the coldest part of the year. Seasonal frosts of N_2 then appear in the polar regions, but most of the N_2 inventory remains in Sputnik Planitia. The lower the TI, albedo or emissivity of N_2 ice, the more mobile the ice. This leads to larger and more time-resistant seasonal frosts. When TI is too low for N_2 ice to remain in the crater, N_2 forms only seasonal polar caps. However, the high-TI simulations are more consistent with the observations, not only because they form a thick N_2 glacier in Sputnik Planitia, but also because they predict a realistic evolution of the surface pressure (Fig. 2a and Extended Data Fig. 2).

Assuming realistic N_2 ice albedos in our model, we could obtain a threefold increase in surface pressure between 1988 and 2015, as observed with stellar occultations^{4–6}, reaching about 1.1 Pa when the New Horizons radio occultation experiment measured it in 2015¹⁵. The pressure increased during this period because the sub-solar point was then at the latitudes of Sputnik Planitia and the N_2 ice insolation was

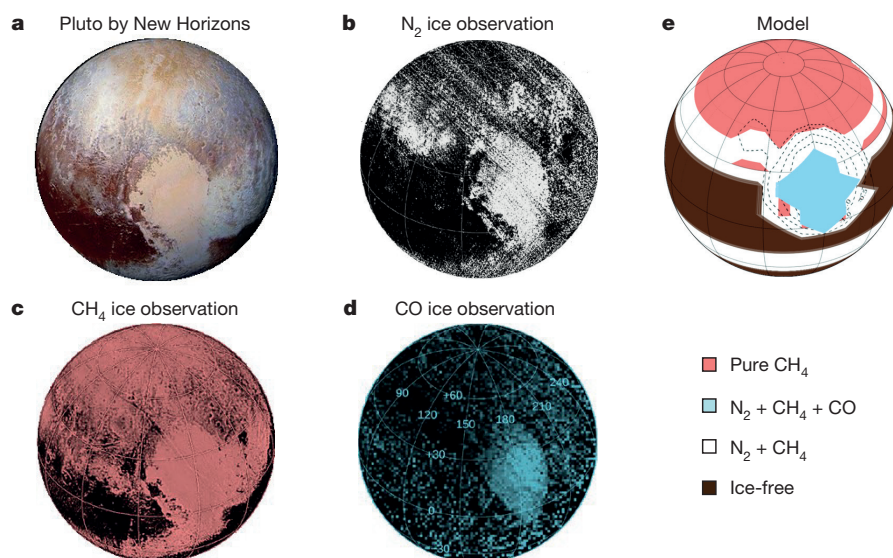


Figure 3 | Comparison between the observed^{1,2} and modelled ice distributions in July 2015 with a slightly brighter and colder methane frost than in Figs 1 and 2. a–d, Observed; e, modelled. $TI = 800 \text{ J s}^{-1/2} \text{ m}^{-2} \text{ K}^{-1}$, $A_{\text{CH}_4} = 0.7$ and $\varepsilon_{\text{CH}_4} = 0.95$. This scenario allows N_2 frost to condense at mid-latitudes as observed.

near maximum. After 2015, the modelled mean pressure decreased because the insolation in Sputnik Planitia was reduced, initially because the subsolar point was at higher latitudes and later because Pluto moved away from the Sun.

The stellar occultation pressure observations provide quantitative constraints for models. If we tune the N_2 ice albedo to obtain surface pressure in the 1–1.2 Pa range in 2015 as observed (Fig. 2), we find that low TI leads to a stronger increase of surface pressures in 1988–2015. In our simulations, the ratio of surface pressure P_s in 2015 to that in 1988—that is, $P_s(2015)/P_s(1988)$ —is near 2.5, 3, 4.5 and 6 for TI values of $1,000 \text{ J s}^{-1/2} \text{ m}^{-2} \text{ K}^{-1}$, $800 \text{ J s}^{-1/2} \text{ m}^{-2} \text{ K}^{-1}$, $600 \text{ J s}^{-1/2} \text{ m}^{-2} \text{ K}^{-1}$ and $400 \text{ J s}^{-1/2} \text{ m}^{-2} \text{ K}^{-1}$ respectively. TI values lower than $500 \text{ J s}^{-1/2} \text{ m}^{-2} \text{ K}^{-1}$ are thus ruled out. The higher values of TI are sufficient to maintain a substantial surface pressure of several millipascals throughout Pluto's orbit (Fig. 2).

Within that context, where might CO be expected to condense on Pluto? Our simulations show that in the moderate- TI to high- TI cases CO accumulates with N_2 ice in Sputnik Planitia and never forms pure CO deposits, in agreement with the New Horizons detections of CO in Sputnik Planitia² (Fig. 3d). Although CO is almost as volatile as N_2 , it is found in trace amounts in the atmosphere (thus requiring very low temperatures to condense) as a consequence of its low mixing ratio in the surface N_2 :CO ice mixture and Raoult's law, controlling the vapour pressure equilibrium. We do not independently calculate the N_2 :CO surface ice mixing ratio, but rather impose 0.3% as derived from telescopic observations¹⁶. Figure 2b shows the evolution of atmospheric CO in the reference simulation with $TI = 800 \text{ J s}^{-1/2} \text{ m}^{-2} \text{ K}^{-1}$. The mean gas volume mixing ratio remains close to 0.04%, in good agreement with the telescopic measurements made in 2010¹⁷. In contrast, for the simulations with low TI ($TI < 200 \text{ J s}^{-1/2} \text{ m}^{-2} \text{ K}^{-1}$) and no permanent N_2 ice deposits, the CO gas mixing ratio reaches 20%–30% in 2010, which is unrealistic and confirms the presence of a permanent cold trap in Sputnik Planitia.

The modelled methane cycle is fundamentally different. Like CO, CH_4 tends to accumulate on N_2 ice in Sputnik Planitia, which acts as a cold trap. However, CH_4 is much less volatile than CO and after 50,000 years the methane cycle forms seasonal CH_4 frosts through vapour–pressure interaction with atmospheric methane. Seasonal methane frosts are present in both hemispheres during autumn, winter and spring, excluding an equatorial belt which always remains ice-free (Fig. 1 and Extended Data Fig. 3). These seasonal deposits explain the bright northern pole observed in 1985–1990⁷, 1994^{8,18} and

2002¹⁸ and the methane-covered northern hemisphere seen by New Horizons.

Figure 2b shows the evolution of atmospheric methane in our reference simulation. The two seasonal peaks are induced by the sublimation of the seasonal polar frost during northern and southern spring. The quantities are very sensitive to model parameters, but using realistic CH_4 ice albedo and emissivity values ($A_{\text{CH}_4} = 0.6$ and $\varepsilon_{\text{CH}_4} = 0.9$) the mean volume mixing ratio reaches 0.4%–0.5% in 2015 in agreement with the observations^{17,19}. The ratio reaches more than 2% in the opposite season, because the atmosphere is then ten times thinner and provides less dilution of methane.

We discovered that a higher CH_4 ice albedo ($A_{\text{CH}_4} > 0.65$) enables the ice to be sufficiently cold in autumn and winter to allow N_2 to condense and form a latitudinal band around 25° – 60°N (Fig. 3e). This corresponds to the latitudes where N_2 frost was detected outside Sputnik Planitia² (Fig. 3a, b). The frost can last until late spring in both hemispheres (that is, after 2015 in northern spring) and is almost devoid of CO ices. This is the scenario that best explains the zonal band of N_2 frost observed by New Horizons.

In all of our simulations that produce a permanent Sputnik Planitia N_2 glacier, methane is slowly trapped in Sputnik Planitia, from where it can barely sublime. The model ultimately converges towards a methane-poor Pluto atmosphere with no methane frost. In reality, however, some processes must refill the system with methane. For instance, cold-trapped methane may be released when N_2 ice retreats at the edge of the Sputnik Planitia glacier, especially when the obliquity or the perihelion season vary²⁰. Perennial methane deposits may form locally as a result of processes not included in our model²¹, such as the reduced insolation on local slopes or atmospheric adiabatic cooling inducing methane precipitation on mountains.

The very high model sensitivity encountered during our studies and the complexity of the methane ice-rich terrains observed by New Horizons show that many details of the CH_4 cycle remain to be explored. Meanwhile our volatile model provides a first-order explanation for the distribution of ices, the evolution of surface pressure and the atmospheric volatile budget observed on Pluto. A robust prediction from many of our simulations is the disappearance of the mid- and high-latitude frosts currently observed in the northern hemisphere (Fig. 1) and the decrease of pressure and atmospheric methane in the next decade (Fig. 2). Such changes would be detectable telescopically, allowing future observers to test the model.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 June; accepted 19 July 2016.

Published online 19 September; corrected online 30 November 2016

(see full-text HTML version for details).

1. Stern, A. S. *et al.* The Pluto system: initial results from its exploration by New Horizons. *Science* **350**, aad1815 (2015).
2. Grundy, W. M. *et al.* Surface compositions across Pluto and Charon. *Science* **351**, aad9189 (2016).
3. Moore, J. F. *et al.* The geology of Pluto and Charon through the eyes of New Horizons. *Science* **351**, 1284–1293 (2016).
4. Elliot, J. L. *et al.* Changes in Pluto's atmosphere: 1988–2006. *Astrophys. J.* **134**, 1–13 (2007).
5. Olkin, C. B. *et al.* Evidence that Pluto's atmosphere does not collapse from occultations including the 2013 May 04 event. *Icarus* **246**, 220–225 (2015).
6. Sicardy, B. *et al.* Pluto's atmosphere from the 29 June 2015 ground-based stellar occultation at the time of the New Horizons flyby. *Astrophys. J. Lett.* **819**, L38 (2016).
7. Buie, M. W., Tholen, D. J. & Horne, K. Albedo maps of Pluto and Charon—initial mutual event results. *Icarus* **97**, 211–227 (1992).
8. Stern, S. A., Buie, M. W. & Trafton, L. M. HST high-resolution images and maps of Pluto. *Astron. J.* **113**, 827–843 (1997).
9. Forget, F., Bertrand, T., Vangvichith, M. & Leconte, A. 3D global climate model of the Pluto atmosphere to interpret New Horizons observations, including the N₂, CH₄ and CO cycles and the formation of organic hazes. *Abstract DPS* **47**, 105.12, <http://adsabs.harvard.edu/abs/2015DPS...4710512F> (2015).
10. Lellouch, E., Stansberry, J., Emery, J., Grundy, W. & Cruikshank, D. Thermal properties of Pluto's and Charon's surfaces from Spitzer observations. *Icarus* **214**, 701–716 (2011).
11. Hansen, C. J. & Paige, D. A. Seasonal nitrogen cycles on Pluto. *Icarus* **120**, 247–265 (1996).
12. Young, L. A. Pluto's seasons: new predictions for New Horizons. *Astrophys. J.* **766**, L22 (2013).
13. Hamilton, D. P. *et al.* The rapid formation of Sputnik Planitia early in Pluto's history. *Nature* <http://dx.doi.org/10.1038/nature20586> (2016).
14. James, P. B., Kieffer, H. H. & Paige, D. A. in *Mars* (eds Kieffer, H., Jakosky, B., Snyder C. & Matthews, M.) 934–968 (Univ. Arizona Press, 1992).
15. Gladstone, G. R. *et al.* The atmosphere of Pluto as observed by New Horizons. *Science* **351**, aad8866 (2016).
16. Merlin, F. New constraints on the surface of Pluto. *Astron. Astrophys.* **582**, A39 (2015).
17. Lellouch, E., de Bergh, C., Sicardy, B., Kauf, H. U. & Smette, A. High resolution spectroscopy of Pluto's atmosphere: detection of the 2.3 μm CH₄ bands and evidence for carbon monoxide. *Astron. Astrophys.* **530**, L4 (2011).
18. Buie, M. W., Grundy, W. M., Young, E. F., Young, L. A. & Stern, S. A. Pluto and Charon with the Hubble Space Telescope. II. Resolving changes on Pluto's surface and a map for Charon. *Astron. J.* **139**, 1128–1143 (2010).
19. Lellouch, E. *et al.* Exploring the spatial, temporal, and vertical distribution of methane in Pluto's atmosphere. *Icarus* **246**, 268–278 (2015).
20. Earle, A. M. & Binzel, R. P. Pluto's insolation history: latitudinal variations and effects on atmospheric pressure. *Icarus* **250**, 405–412 (2015).
21. Stansberry, J. A. *et al.* A model for the overabundance of methane in the atmospheres of Pluto and Triton. *Planet. Space Sci.* **44**, 1051–1063 (1996).

Acknowledgements We thank the New Horizons team for this successful mission. We also thank E. Lellouch, E. Millour and M. J. Wolff for comments on the manuscript. We are grateful to M. Vangvichith for her contribution to an early version of the model, and to the Institut de Formation Doctorale for supporting this work.

Author Contributions F.F. and T.B. designed and developed the model. T.B. performed the simulations. Both authors contributed to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.F. (forget@lmd.jussieu.fr).

METHODS

Pluto's surface is represented by a grid of 32 longitudes \times 24 latitudes. At each point the model computes the surface radiative budget, the exchange of heat with the subsurface by conduction, and the exchanges of volatile (N_2 , CO , CH_4) with the atmosphere.

Insolation and radiative cooling. At each time step the local solar insolation is calculated taking into account the variation of the Pluto–Sun distance throughout its orbit, the seasonal inclination and the diurnal cycle. The atmosphere is assumed to be transparent at all wavelengths.

Thermal conduction into the subsurface. The heat flux from and to the subsurface is computed using a classical heat conduction model with thermal inertia I as the key parameter controlling the influence of the subsurface heat storage and conduction on the surface temperature: $I = \sqrt{\lambda C}$, with λ the heat conductivity of the ground (in units of $J s^{-1} m^{-1} K^{-1}$) and C the ground volumetric specific heat ($J m^{-3} K^{-1}$). In practice, we thus use I as the key model parameter, assuming a constant value for $C = 10^6 J m^{-3} K^{-1}$ and making λ vary accordingly. On Pluto one needs to simultaneously capture (1) the short-period diurnal thermal waves in the near-surface, low-TI terrain and (2) the much longer seasonal thermal waves which can penetrate deep in the high-TI substrate. The diurnal thermal inertia is set to $20 J s^{-1/2} m^{-2} K^{-1}$, as inferred from Spitzer thermal observations¹⁰, and the seasonal TI is varied in the range $200\text{--}2,000 J s^{-1/2} m^{-2} K^{-1}$, as discussed in the main text.

The modelled diurnal and annual skin depths are 0.008 m and 10–100 m respectively. To adequately resolve these scale lengths, the subsurface is divided into 22 discrete layers, with a geometrically stretched distribution of layers with higher resolution near the surface (the depth of the first layer is 1.4×10^{-4} m) and a coarser grid for deeper layers (the deepest layer depth is near 300 m). Our simulations are performed assuming no internal heat flux.

Volatile condensation and sublimation. We computed the phase changes of N_2 , CH_4 and CO using the thermodynamic relations from ref. 22, taking into account the α - and β -phase transitions. The rate of N_2 condensation or sublimation is derived from the amount of latent heat required to keep temperatures at the frost point if N_2 ice is present. CH_4 and CO are minor constituents and their exchange with the atmosphere depends on the turbulent fluxes given by:

$$F = \rho C_d |U| (q - q_{\text{surf}}) \quad (1)$$

with q_{surf} the saturation vapour pressure mass mixing ratio (in $kg kg^{-1}$) at the considered surface temperature, q the mass mixing ratio in the atmosphere, ρ the air density, U the horizontal wind velocity at $z_1 = 5$ m above the surface ($0.5 m s^{-1}$), and C_d the drag coefficient at 5 m above the local surface set to 0.06.

Since the ices may form solid solutions, to compute q_{surf} when N_2 ice is present we applied Raoult's law considering the mixtures $N_2:CH_4$ and $N_2:CO$ with 0.5% of methane and 0.3% of CO respectively, as retrieved from telescopic observations¹⁶.

The modelled volatile cycles are slightly affected by the limited grid resolution. For instance, in Fig. 2b the 'saw-tooth' variations around $L_s = 200^\circ$ are artefacts resulting from the sudden sublimation of latitudinal bands of methane ice on the low-resolution grid.

Atmospheric horizontal 'transport' and effect of topography. To represent horizontal transport after surface exchanges, atmospheric species are mixed with an adjustable intensity, taking into account the effect of topography. For N_2 , at each time step the surface pressure p_s at altitude z is forced towards the global mean value by integrating the following equation:

$$\frac{\partial p_s}{\partial t} = -\frac{1}{\tau_{N_2}} \left(p_s - \langle p_s \rangle \frac{e^{-z/H}}{\langle e^{-z/H} \rangle} \right) \quad (2)$$

with H the near-surface scale height (about 18 km), $\langle \rangle$ the planetary average operator, and τ_{N_2} the characteristic mixing timescale (in seconds). Similarly, for a trace species like CH_4 , the local mass mixing ratio q mixing is given by:

$$\frac{\partial q}{\partial t} = -\frac{1}{\tau_{CH_4}} \left(q - \frac{\langle q p_s \rangle}{\langle p_s \rangle} \right) \quad (3)$$

On the basis of tests performed with the Pluto General Circulation Model⁹ (see below), we set $\tau_{CH_4} = 10^7$ s (about four terrestrial months), and $\tau_{N_2} = 1$ s (instantaneous mixing). Note that adiabatic cooling inducing atmospheric condensation is not taken into account.

Glacier flow modelling. In the modelled Sputnik Planitia basin, N_2 tends to condense first where there is N_2 ice, essentially because the albedo difference between the ice and the substrate favours further condensation on the ice. Consequently, the ice tends to accumulate in only a fraction of the basin. In reality, massive N_2 ice deposits should flow like terrestrial glaciers³. This can affect the evolution of surface pressure, since a smaller surface of N_2 ice will be available for condensation and sublimation. To address this issue, we model the flow of N_2 ice in the basin by redistributing it locally from a point with a large amount of ice to the nearest-neighbour points with less ice. We use a characteristic speed of 7 cm per Pluto day (about 1 cm per terrestrial day), which allows the crater to be filled with ice after two Pluto years.

The three-dimensional GCM used to tune the volatile transport model. The full three-dimensional GCM⁹ will be described in detail elsewhere²³. It is a three-dimensional atmospheric model which uses the same horizontal grid as does the two-dimensional volatile transport model used here, but with 25 layers in the vertical (starting at height 7 m, and reaching 250 km, with most of the levels in the first 15 km in order to obtain a good resolution close to the surface). In addition to the subsurface thermal conduction and surface volatile condensation and sublimation processes already used in the volatile-transport model described above, the following features are included:

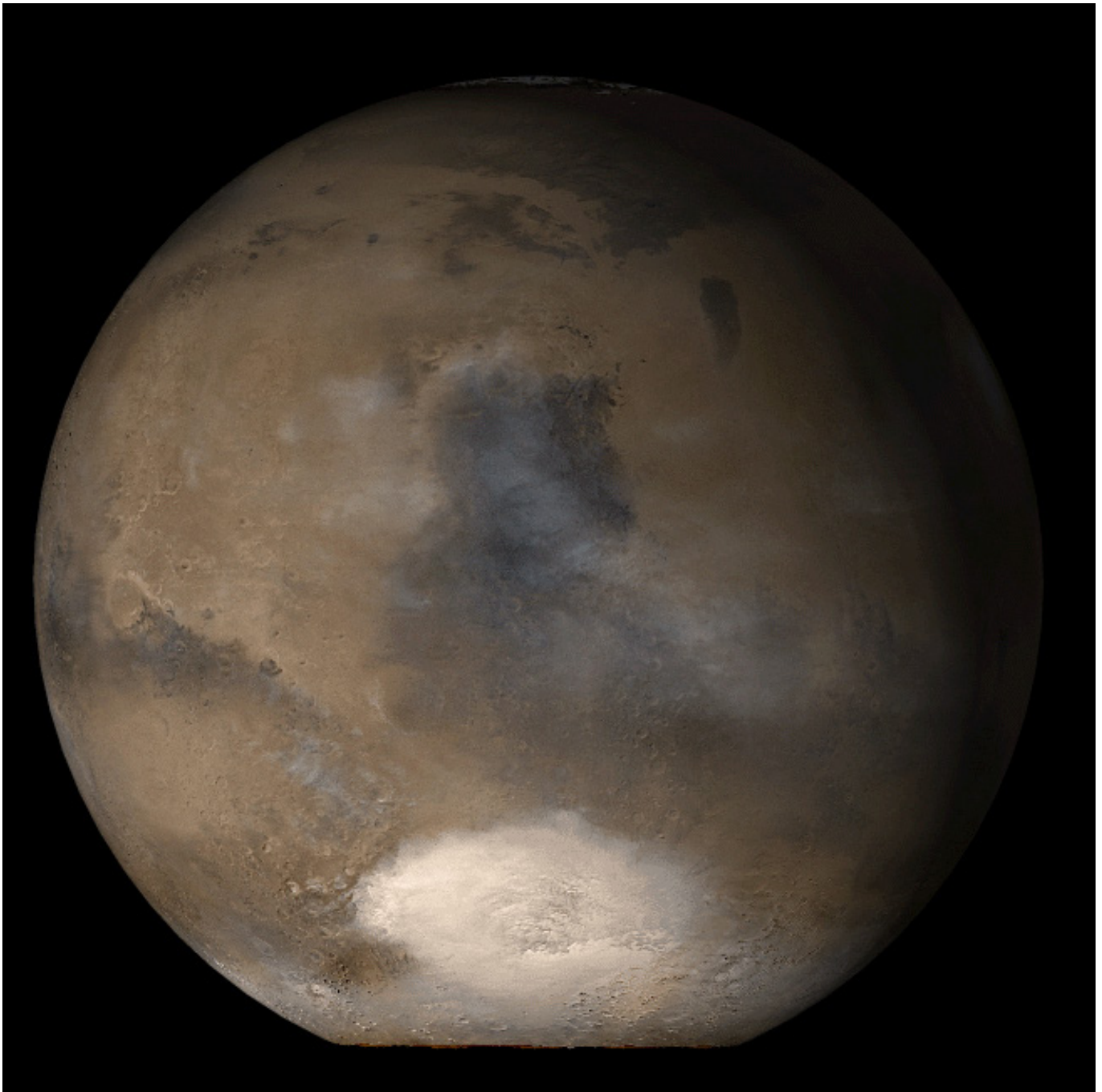
- (1) A finite difference three-dimensional dynamical core to solve the primitive equation of meteorology.
- (2) A three-dimensional transport scheme to transport atmospheric species like CO gas and CH_4 gas and ice.
- (3) In the atmosphere, radiative heating and cooling is taken into account by computing the amount of CH_4 molecules, and cooling by the thermal infrared rotational lines of CO is taken into account by assuming a constant mixing ratio of 0.05%.
- (4) Turbulent mixing and convection in the boundary layer.
- (5) Atmospheric molecular thermal conduction and viscosity.
- (6) Atmospheric CH_4 condensation and transport of CH_4 ice cloud particles.

Design of the reference simulations and comparison with observations. The simulations shown in Fig. 2a have been performed using an N_2 ice emissivity of 0.8 and different N_2 ice albedo chosen to yield realistic pressure in 2015: $A_{N_2} = 0.73$ for $TI = 500 J s^{-1/2} m^{-2} K^{-1}$; $A_{N_2} = 0.71$ for $TI = 800 J s^{-1/2} m^{-2} K^{-1}$; $A_{N_2} = 0.69$ for $TI = 1,500 J s^{-1/2} m^{-2} K^{-1}$.

The observed CH_4 mixing ratios shown in Fig. 2b (error bars) are in the range $0.38 \pm 0.06\%$ (for 2008) and $0.34 \pm 0.06\%$ (for 2012)¹⁹ (E. Lellouch, personal communication), and CO mixing ratios are in the range $0.05^{+0.1}_{-0.025}\%$ (for 2010)¹⁷ with 2σ error bars. The central values of the CH_4 mixing ratios are computed from the weighted average of the values obtained from the different simulations in Lellouch *et al.*¹⁹, by their own uncertainty. The error bar takes into account the dispersion between the central values and the noise uncertainty.

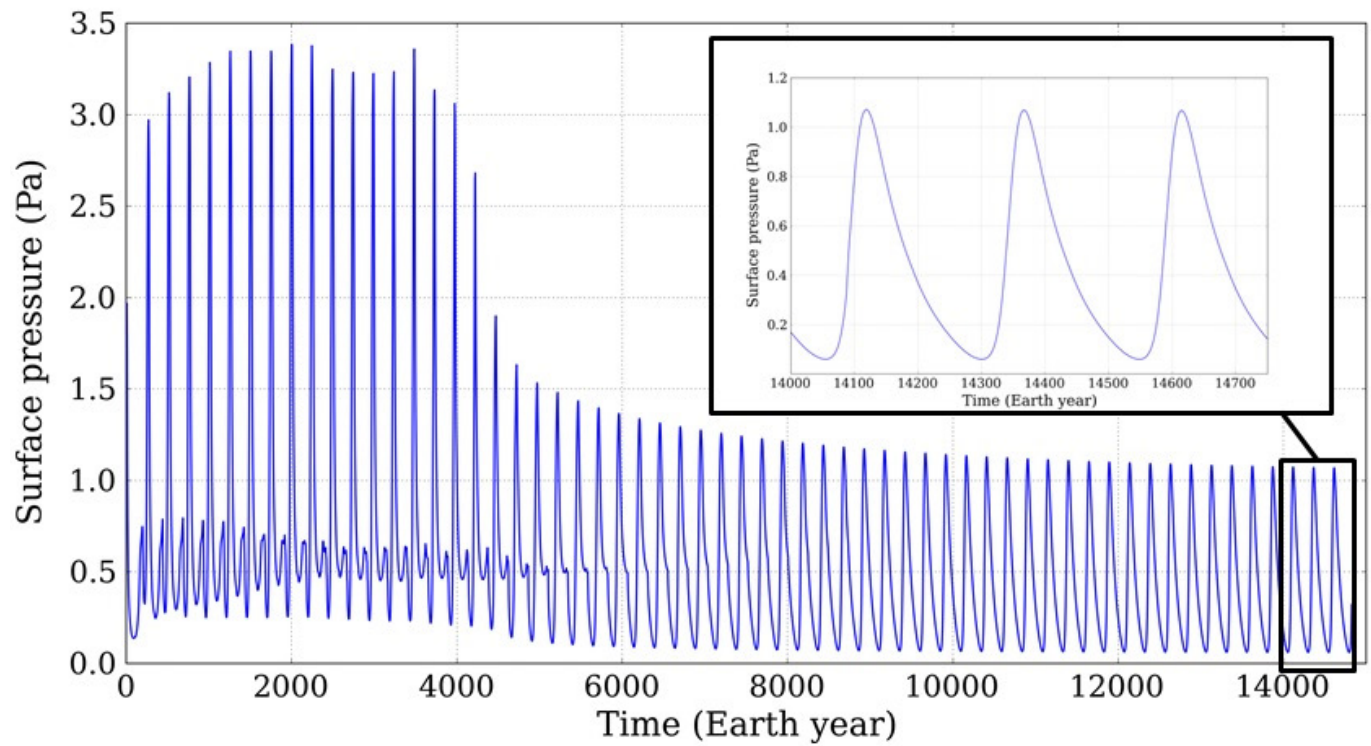
Code availability. All model versions are freely available upon request by contacting F.F. (forget@lmd.jussieu.fr).

22. Fray, N. & Schmitt, B. Sublimation of ices of astrophysical interest: a bibliographic review. *Planet. Space Sci.* **57**, 2053–2080 (2009).
23. Forget, F. *et al.* A post-New Horizons global climate model of Pluto including the N_2 , CH_4 and CO cycles. *Icarus* (submitted).

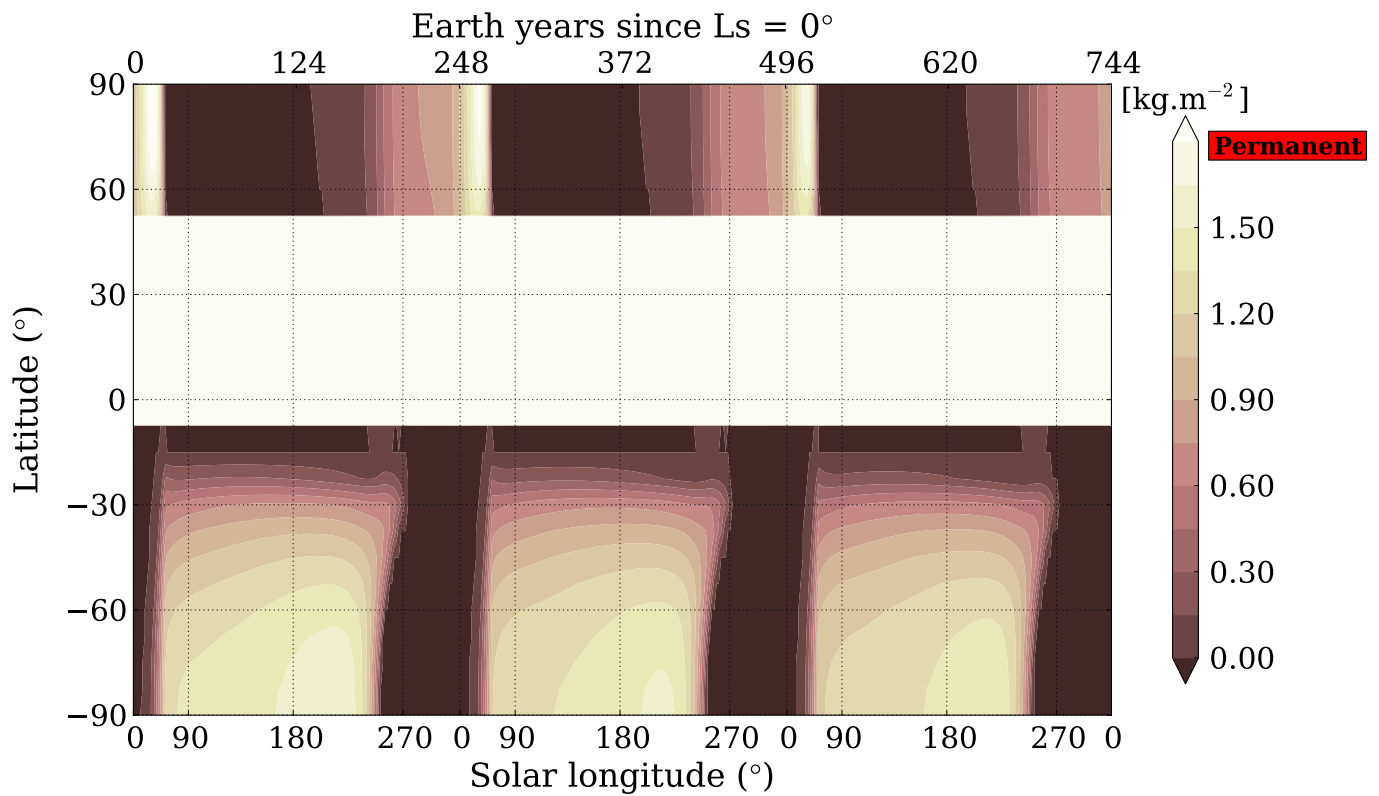


Extended Data Figure 1 | Mars in mid-southern spring. This composite image shows the 7-km-deep, 2,000-km-wide deep Hellas Planitia impact basin covered by seasonal CO₂ ice. The physical processes that favour the condensation of the CO₂ atmosphere at the bottom of the Hellas basin is the same as for Pluto's N₂ atmosphere in Sputnik Planitia. In both basins, surface pressures are larger than in the surrounding regions and condensation temperatures T_c are higher. This corresponds to a stronger

thermal infrared cooling (proportional to T_c^4), which is balanced by a proportionally more intense condensation rate (latent heat). Images obtained by the Mars Global Surveyor Mars Orbiter Camera in November 2006, at solar longitude (the Mars–Sun angle, measured from the Northern Hemisphere spring equinox where $L_s = 0^\circ$) $L_s = 137^\circ$ (image credit NASA/JPL/MSSS; <http://photojournal.jpl.nasa.gov/catalog/PIA01888>).



Extended Data Figure 2 | Evolution of surface pressure during the first 15,000 Earth years of the reference simulation (same as Fig. 1). At least 10,000 Earth years are necessary to equilibrate the N_2 ice in the basin and obtain a year-to-year repetitive pattern of surface pressure.



Extended Data Figure 3 | Latitudinal distribution of CH₄ ice over three Pluto years in our reference simulation, as in Figs 1 and 2. The colour scale measures methane ice on the surface, in kilograms per square metre. The time axis starts at the solar longitude $L_s = 0^\circ$. In addition to the permanent methane ice reservoir in the Sputnik Planitia basin, the methane cycle exhibits seasonal polar frosts of methane. The southern

polar frosts remain in place longer than the northern frost, owing to the longer duration of the southern winter season. At mid-northern latitudes, the Sputnik Planitia basin is a permanent reservoir of CH₄ ice. CH₄ frost preferentially accumulates at the poles during autumn and winter, but better resists sublimation at low latitudes during late spring, especially in the southern hemisphere.

Reorientation and faulting of Pluto due to volatile loading within Sputnik Planitia

James T. Keane¹, Isamu Matsuyama¹, Shunichi Kamata² & Jordan K. Steckloff^{3,4}

Pluto is an astoundingly diverse, geologically dynamic world. The dominant feature is Sputnik Planitia—a tear-drop-shaped topographic depression approximately 1,000 kilometres in diameter possibly representing an ancient impact basin^{1,2}. The interior of Sputnik Planitia is characterized by a smooth, craterless plain three to four kilometres beneath the surrounding rugged uplands, and represents the surface of a massive unit of actively convecting volatile ices (N₂, CH₄ and CO) several kilometres thick^{1–5}. This large feature is very near the Pluto–Charon tidal axis. Here we report that the location of Sputnik Planitia is the natural consequence of the sequestration of volatile ices within the basin and the resulting reorientation (true polar wander) of Pluto. Loading of volatile ices within a basin the size of Sputnik Planitia can substantially alter Pluto's inertia tensor, resulting in a reorientation of the dwarf planet of around 60 degrees with respect to the rotational and tidal axes. The combination of this reorientation, loading and global expansion due to the freezing of a possible subsurface ocean generates stresses within the planet's lithosphere, resulting in a global network of extensional faults that closely replicate the observed fault networks on Pluto. Sputnik Planitia probably formed northwest of its present location, and was loaded with volatiles over million-year timescales as a result of volatile transport cycles on Pluto^{6,7}. Pluto's past, present and future orientation is controlled by feedbacks between volatile sublimation and condensation, changing insolation conditions and Pluto's interior structure.

Centred at 176° E, 24° N, Sputnik Planitia is close to the Pluto–Charon tidal axis (Fig. 1). The alignment of large geologic features with the principal axes of inertia is the hallmark of true polar wander⁸ (TPW). In a minimum energy state, planets align their minimum (maximum) principal axis of inertia with the tidal (spin) axis. TPW occurs when mass is redistributed within the planet and the geographic locations of these axes change. To remain in a minimum energy state, the planet reorients to realign these principal axes with the tidal/spin axes. Notable examples of planetary TPW include: the reorientation of Enceladus to place the plume-producing tiger stripes at the south pole^{9,10}; the reorientation of Mars to place the Tharsis volcanic rise at the equator¹¹; and the reorientation of the Moon to place the South Pole–Aitken impact basin near the south pole¹² (see ref. 8 for a review). How a planet reorients is controlled by the feature's mass anomaly, Q' , which is quantified in terms of degree-2 gravity coefficients: $Q' = -J_2^{SP*}/J_2^{RF}$ (where J_2^{SP*} is the degree-2 zonal spherical harmonic gravity coefficient of Sputnik Planitia when centred at Pluto's north pole and J_2^{RF} is the degree-2 zonal spherical harmonic gravity coefficient of Pluto's remnant figure; Methods). In the absence of a remnant figure, positive (negative) mass anomalies will reorient the planet to align with the tidal (spin) axis. The proximity of Sputnik Planitia to Pluto's tidal axis suggests that it is a positive mass anomaly.

TPW is counteracted by the planet's non-hydrostatic, elastically supported tidal-rotational bulge, which can preserve a previous tidal-rotational potential—a remnant figure. At present, no bulge (remnant

or otherwise) has been observed at Pluto¹³. In the absence of measurements of Pluto's figure, we constructed a four-layer model of Pluto using viscoelastic Love number theory¹⁴, using Pluto's mass and radius as constraints. Our nominal Pluto model consists of a silicate-rich core and a liquid water ocean overlaid by an H₂O-rich weak crust and elastic lithosphere (Methods).

Using our Pluto model, we evaluate how Pluto would reorient in response to the formation of Sputnik Planitia. The mass anomaly of Sputnik Planitia depends on its presumed density structure. Although topographic depressions are negative mass anomalies, impact basins (as Sputnik Planitia is hypothesized to be on the basis of its quasi-circular shape²) have stochastic mass anomalies owing to the complicated density structures that formed during the impact process^{12,15} (Extended Data Fig. 1). Unavoidable impact basin components, such as ejecta blankets, can substantially offset the negative mass anomaly of the topographic depression (Extended Data Figs 2, 3). We remain agnostic as to the true structure of the Sputnik Planitia basin, and provide relationships that link Q' to the thickness of ice within the basin for several different simple basin structures (Fig. 2b–e, Extended Data Figs 1–3, Methods).

With these assumptions, we determined the possible initial locations of Sputnik Planitia as a function of Q' . We self-consistently adjusted Q' to account for the flexural response of Pluto's elastic lithosphere,

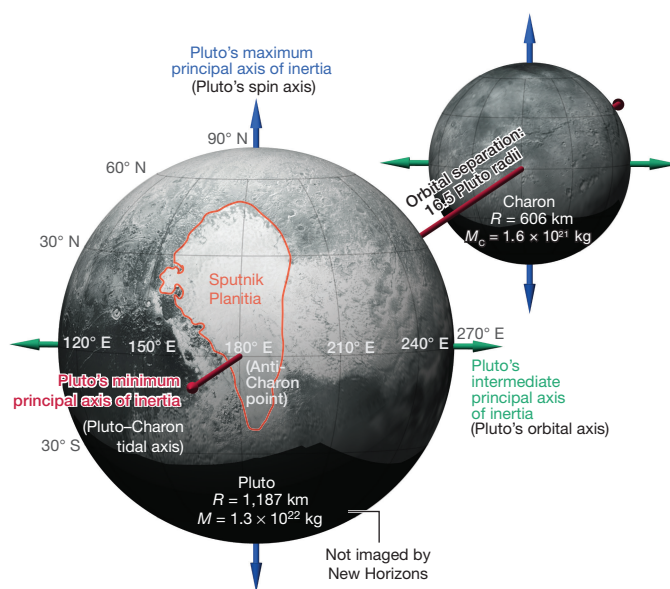


Figure 1 | Geometry of Sputnik Planitia in the Pluto–Charon system. Orthographic spherical projections of Pluto and Charon (to scale), with Sputnik Planitia and the principal axes of inertia labelled. Base map: NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute.

¹Lunar and Planetary Laboratory, Department of Planetary Science, University of Arizona, Tucson, Arizona 85721, USA. ²Creative Research Institution, Hokkaido University, Sapporo, Japan.

³Purdue University, Department of Earth, Atmospheric, and Planetary Sciences, West Lafayette, Indiana 47907, USA. ⁴Planetary Science Institute, Tucson, Arizona 85719, USA.

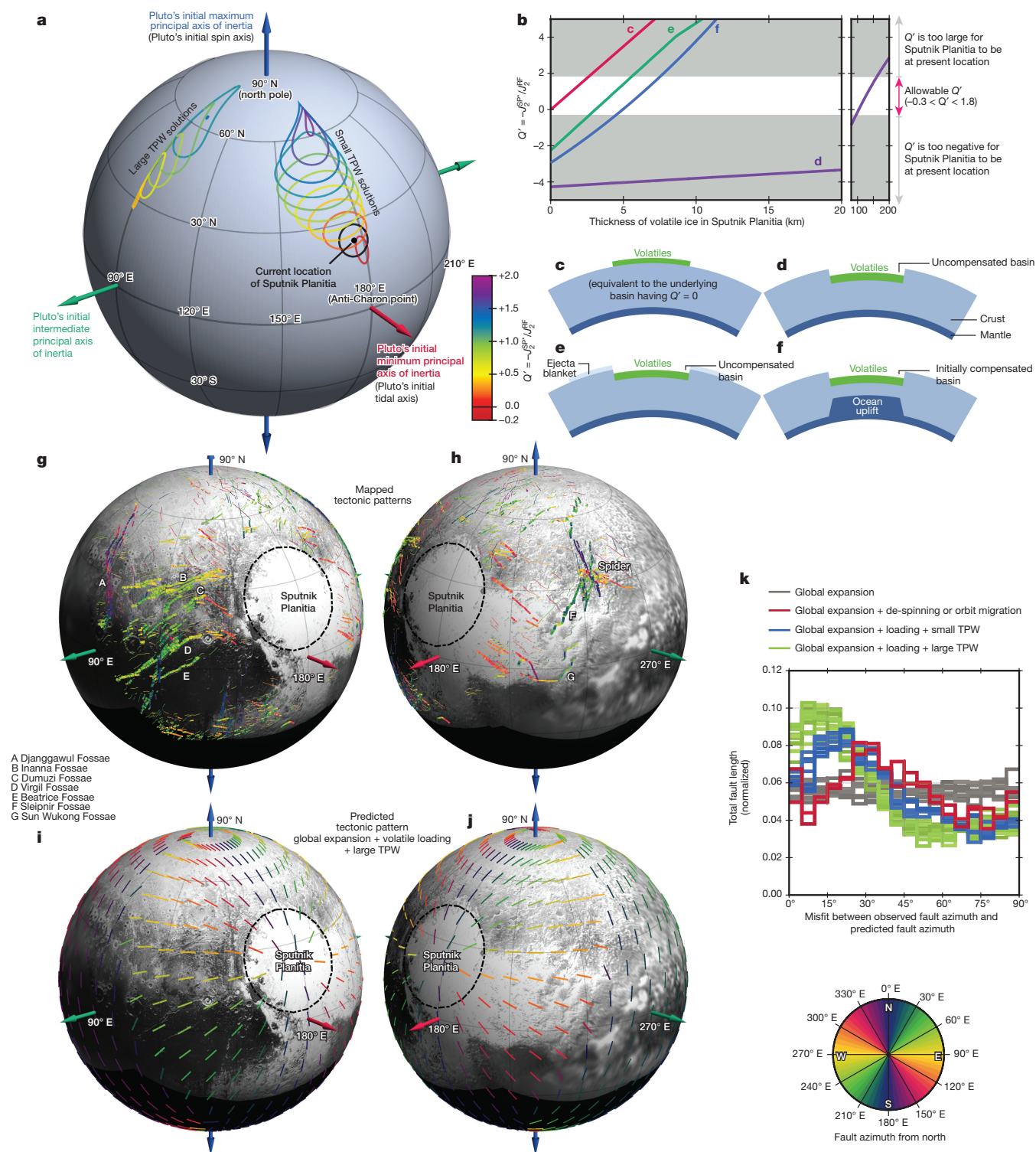


Figure 2 | True polar wander solutions for Sputnik Planitia and the resulting tectonic patterns. **a**, The possible initial locations of Sputnik Planitia before reorientation with respect to the principal axes of the remnant figure, as a function of Q' . **b–f**, Q' as a function of volatile thickness (**b**) for four idealized models of Sputnik Planitia (**c–f**). **c**, Volatiles on the surface (equivalent to volatiles loading a basin with $Q' = 0$). **d**, Volatiles filling an uncompensated basin. **e**, Volatiles filling an uncompensated impact basin surrounded by an ejecta blanket. **f**, Volatiles filling an impact basin that was initially compensated by an uplift in the subsurface ocean¹⁶. **g, h**, Tectonic features mapped in

publicly available New Horizons imagery. Base map: NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute. Faults are coloured by azimuth, and the line width corresponds to our certainty in the nature of fault: thick (thin) lines are unambiguously (possible) faults. **i, j**, Best-fitting predicted tectonic pattern resulting from global expansion, loading and a large TPW event. **k**, Binned misfit between observed and predicted fault azimuths (following the procedure used in ref. 19). Each line represents a different tectonic model (Extended Data Fig. 5; Methods). Large TPW solutions provide the best match to the observed distribution of faults.

which effectively increases the ice thickness that is required to produce a given reorientation. We performed a parameter space search—placing models of Sputnik Planitia with varying Q' values across the entire surface of Pluto and evaluating how Pluto reoriented in response. Contours in Fig. 2a bound the initial locations of Sputnik Planitia (as a function of Q') that reoriented Pluto to place Sputnik Planitia within 5° of its present location. Several key observations can be made from this figure. First, Sputnik Planitia could not have formed in any random location; the initial positions are limited to regions of a single quadrant of the northern, anti-Charon side of Pluto. This arises from energy constraints during single-episode TPW that prohibit the perturbing anomaly from crossing latitude and longitude lines of the remnant figure's principal axes. Within this quadrant, the available initial conditions are further constrained, as Sputnik Planitia is not exactly at the tidal axis. Second, the parameter space overwhelmingly favours positive Q' values (Extended Data Fig. 4). Q' is constrained to be between -0.3 and 1.8 . This reveals that Sputnik Planitia is not a purely uncompensated topographic depression, as such a feature would have $Q' = -4$ (Fig. 2b, d), which would completely overwhelm Pluto's remnant figure and reorient Sputnik Planitia to the north pole. Ices alone are not sufficient to cancel out this negative anomaly, as it would require an untenable volume of ice within Sputnik Planitia (>100 km; Fig. 2b, d). Instead, some combination of ice and basin structure (for example, an ejecta blanket (Fig. 2b, e) or uplift of the subsurface ocean¹⁶ (Fig. 2b, f)) is required for Sputnik Planitia to be at its present location. Third, depending on the presumed structure of Sputnik Planitia, the required ice thicknesses range from 0 to 10 km (Fig. 2b–f, Extended Data Fig. 3, Methods), which is

consistent with previous estimates of ice thickness within Sputnik Planitia that are based on inferences of basin geometry², isostasy of mountain-sized H_2O icebergs floating in Sputnik Planitia, and modelling of solid-state convection of ice within Sputnik Planitia^{4,5}. Finally, there are two families of TPW solutions: 'large' solutions, where Sputnik Planitia started northwest of its present location; and 'small' solutions, where Sputnik Planitia started north of its present location (Extended Data Fig. 4a, b). Large TPW solutions occur when the intermediate and minimum principal axes of inertia swap (hence the minimum Q' value for large TPW solutions, which are related to the moment differences).

As a planet reorients, each surface location experiences a change in the tidal rotational potential. This builds stress in the lithosphere, eventually resulting in faults with a characteristic global pattern¹⁷. Pluto possesses a global non-random system of extensional faults^{1,2} (Fig. 2g, h). The lack of observed compressional or translational faults probably reflects global expansion due to the freezing of a subsurface ocean¹⁸. Using our initial locations of Sputnik Planitia (Fig. 2a), we calculated the tectonic patterns for a range of possible TPW scenarios, including the effects of reorientation, global expansion and loading of ice within Sputnik Planitia (Methods). Figure 2i, j shows our best-fitting TPW solution. Proximal to Sputnik Planitia, faults are quasi-radial, primarily from loading stresses. Distal to Sputnik Planitia, TPW stresses dominate and the orientations of the faults change. The predicted tectonic patterns are not strongly sensitive to the bulk properties of Pluto, but are sensitive to the initial location, size and thickness of ice within Sputnik Planitia (Extended Data Fig. 5). We quantified how well our predicted fault geometries fit by

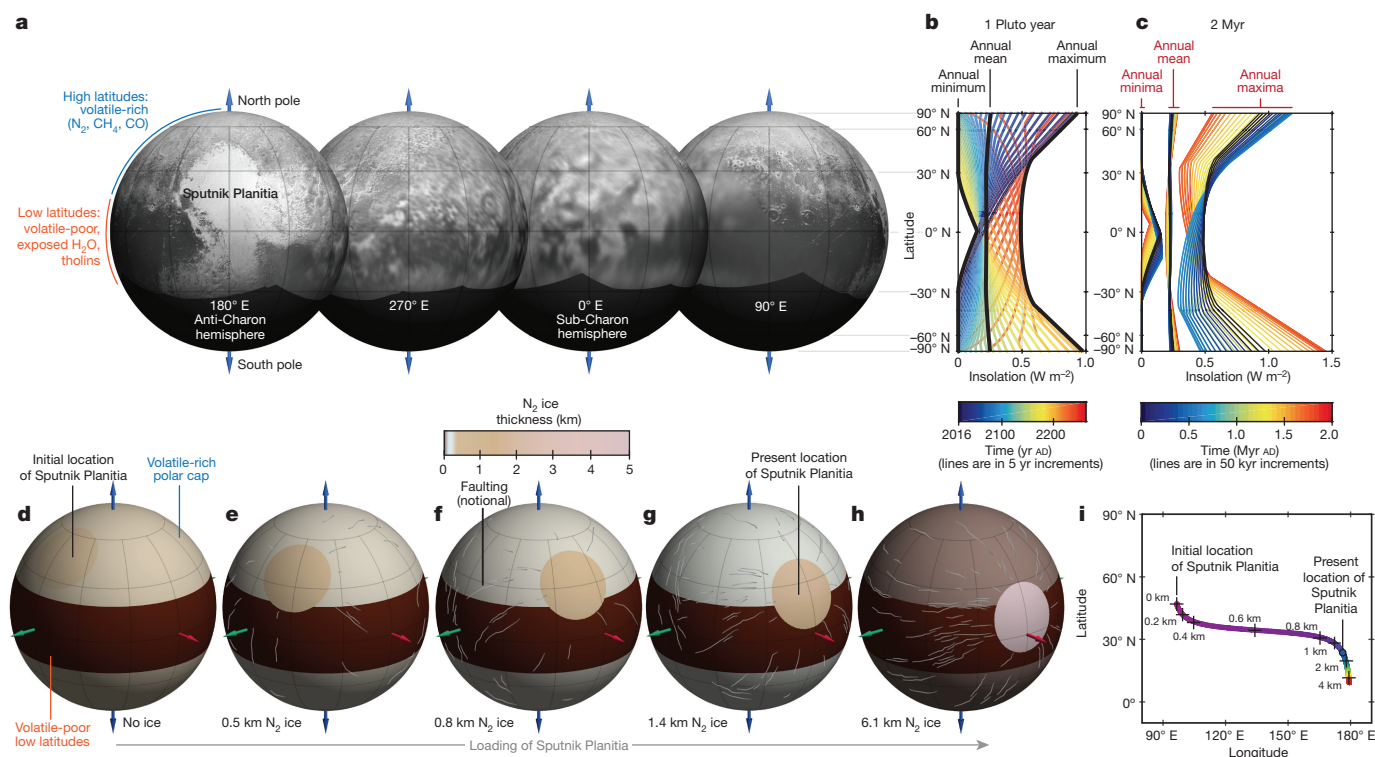


Figure 3 | Insolation patterns and volatile-driven TPW of Pluto.

a, Orthographic spherical projection of Pluto, highlighting the latitudinal trends in albedo and volatile content. Base map: NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute. **b**, Insolation as a function of time on Pluto. Coloured lines denote average insolation over one Pluto day; black lines denote the minimum, maximum and mean over one Pluto year. **c**, Variations in annual minima, maxima and mean insolation over 2 Myr. **d–h**, A simple model of our proposed volatile-induced reorientation of Pluto, where an equivalent global layer of 200 m

of volatile ice is transferred from the poles to Sputnik Planitia (assuming the underlying basin has $Q' = 0$, as in Fig. 2b, c) for various ice loadings from 0 km to 6.1 km in thickness. Sputnik Planitia initially formed at higher latitude, northwest of its present location. As volatiles migrate into Sputnik Planitia from other cold traps, Pluto reorients and fractures (see Extended Data Fig. 6, Supplementary Videos 1–3 and Methods). **i**, The latitude and longitude of Sputnik Planitia are shown as a function of the thickness of ice within the basin. If volatiles continue to migrate into Sputnik Planitia, Pluto will continue to reorient towards the tidal axis.

measuring the difference between the observed and predicted fault azimuths¹⁹ (Fig. 2k). Large TPW solutions yielded the best fit to the observed fault distribution and are marginally better than small TPW solutions. The observed fault geometry is inconsistent with global expansion, de-spinning or orbit migration²⁰ alone (Extended Data Fig. 5). We do not accurately predict the locations of the Sun Wukong Fossae (SWF) and 'spider' fault network east of Sputnik Planitia. Although our model does predict the confluence of eastward- and northward-trending faults at some distance from Sputnik Planitia (as demonstrated by the accurate fitting of faults west of Sputnik Planitia), the SWF and 'spider' are westward of the predicted eastern transition. This difference may be due to oversimplifications in our model geometry of Sputnik Planitia, inhomogeneity in the lithosphere, or faulting during different stress conditions (Extended Data Fig. 6). Nonetheless, TPW, loading and global expansion provide the most comprehensive single explanation for the global pattern of faults on Pluto.

If ice loading within Sputnik Planitia drove TPW, then this suggests a feedback between the planet's volatile cycle and rotational stability. Such a feedback was previously discounted on the basis of Pluto's insolation geometry²¹. Pluto's large obliquity (122°) means that the equator receives less insolation than the poles²² (Fig. 3b, c). Therefore, one might expect volatiles to accumulate at Pluto's equator, building a stabilizing equatorial bulge and inhibiting TPW²¹. However, while Pluto's equator receives less insolation, it is never the coldest part of the planet (Fig. 3b, c). Pluto's poles oscillate seasonally between long polar nights, and if volatiles are sufficiently mobile the poles may still be the preferred sites for volatile deposition^{7,23}. This may explain why volatiles are found preferentially at higher latitudes on Pluto³ (Fig. 3a). Our TPW solutions suggest that Sputnik Planitia formed at higher latitudes in these regions of enhanced seasonal volatile deposition. We posit that Sputnik Planitia is a cold trap (see Methods), and over time it accreted a large fraction of Pluto's volatile reservoir, driving reorientation (Fig. 3d–h). The final location of Sputnik Planitia is set by Pluto's total volatile reservoir, the remnant figure and the feedback between TPW and volatile stability within Sputnik Planitia. The proximity of Sputnik Planitia to the latitude of minimum mean insolation²⁴ (Fig. 3b) may be evidence that a TPW–volatile stability feedback is active. If volatiles migrate into and out of Sputnik Planitia on seasonal timescales, then Pluto may experience small-amplitude wobbles akin to Earth's annual, atmospheric-pressure-driven wobbles²⁵ (see Methods). Similar volatile-driven reorientation feedbacks may be important for continued geologic activity on other large Kuiper belt objects and planetary bodies with large reservoirs of mobile volatiles.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 May; accepted 26 September 2016.

Published online 16 November 2016.

1. Stern, S. A. *et al.* The Pluto system: initial results from its exploration by New Horizons. *Science* **350**, aad1815 (2015).
2. Moore, J. M. *et al.* The geology of Pluto and Charon through the eyes of New Horizons. *Science* **351**, 1284–1293 (2016).
3. Grundy, W. M. *et al.* Surface compositions across Pluto and Charon. *Science* **351**, aad9189 (2016).
4. McKinnon, W. B. *et al.* Convection in a volatile nitrogen-ice-rich layer drives Pluto's geological vigour. *Nature* **534**, 82–85 (2016).

5. Trowbridge, A. J. *et al.* Vigorous convection as the explanation for Pluto's polygonal terrain. *Nature* **534**, 79–81 (2016).
6. Binzel, R. P. Long term variations of a volatile methane reservoir on Pluto. In *21st Lunar Planet. Sci. Conf.* 87–88 (Lunar and Planetary Institute, 1990).
7. Spencer, J. R. *et al.* in *Pluto and Charon* (eds Stern, S. A. & Tholen, D. J.) 435–474 (Univ. Arizona Press, 1997).
8. Matsuyama, I., Nimmo, F. & Mitrovica, J. X. Planetary reorientation. *Annu. Rev. Earth Planet. Sci.* **42**, 605–634 (2014).
9. Nimmo, F. & Pappalardo, R. T. Diapir-induced reorientation of Saturn's moon Enceladus. *Nature* **441**, 614–616 (2006).
10. Collins, G. C. & Goodman, J. C. Enceladus' south polar sea. *Icarus* **189**, 72–82 (2007).
11. Perron, J. T., Mitrovica, J. X., Manga, M., Matsuyama, I. & Richards, M. A. Evidence for an ancient Martian ocean in the topography of deformed shorelines. *Nature* **447**, 840–843 (2007).
12. Keane, J. T. & Matsuyama, I. Evidence for lunar true polar wander and a past low eccentricity, synchronous lunar orbit. *Geophys. Res. Lett.* **41**, 6610–6619 (2014).
13. Nimmo, F. *et al.* Mean radius and shape of Pluto and Charon from New Horizons images. *Icarus* <http://dx.doi.org/10.1016/j.icarus.2016.06.027> (2016).
14. Peltier, W. R. The impulse response of a Maxwell Earth. *Rev. Geophys. Space Phys.* **12**, 649–669 (1974).
15. Melosh, H. J. *et al.* The origin of lunar mascon basins. *Science* **340**, 1552–1555 (2013).
16. Nimmo, F. *et al.* Reorientation of Sputnik Planitia implies a subsurface ocean on Pluto. *Nature* <http://dx.doi.org/10.1038/nature20148> (2016).
17. Matsuyama, I. & Nimmo, F. Tectonic patterns on reoriented and despun planetary bodies. *Icarus* **195**, 459–473 (2008).
18. Hammond, N. P., Barr, A. C. & Parmentier, E. M. Recent tectonic activity on Pluto driven by phase changes in the ice shell. *Geophys. Res. Lett.* **43**, 6775–6782 (2016).
19. Watters, T. R. *et al.* Global thrust faulting on the Moon and the influence of tidal stresses. *Geology* **43**, 851–854 (2015).
20. Matsuyama, I. & Nimmo, F. Pluto's tectonic pattern predictions. In *44th Lunar Planet. Sci. Conf.* 1399 (Lunar and Planetary Institute, 2013).
21. Rubin, D. P. Polar wander on Triton and Pluto due to volatile migration. *Icarus* **163**, 469–478 (2003).
22. Earle, A. M. & Binzel, R. P. Pluto's insolation history: latitudinal variations and effects on atmospheric pressure. *Icarus* **250**, 405–412 (2015).
23. Hansen, C. J., Paige, D. A. & Young, L. A. Pluto's climate modeled with new observational constraints. *Icarus* **246**, 183–191 (2015).
24. Hamilton, D. P. *et al.* The rapid formation of Sputnik Planitia early in Pluto's history. *Nature* <http://dx.doi.org/10.1038/nature20586> (2016).
25. Gross, R. S., Fukumori, I. & Menemenlis, D. Atmospheric and oceanic excitation of the Earth's wobbles during 1980–2000. *J. Geophys. Res.* **108**, 2370 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the New Horizons science team for their many years of work that resulted in the successful flyby of Pluto, and for promptly releasing the public data and published work^{1–4,13,16} that enabled this research. We note that all names of features on Pluto are informal. J.T.K. acknowledges support from the University of Arizona Theoretical Astrophysics Program and NASA Solar System Workings.

Author Contributions J.T.K. identified the link between Sputnik Planitia and the tidal axis, performed true polar wander analyses, mapping and analysis of tectonic features, proposed connection between polar wander and secular volatile transport, was the primary author of the text and created all figures. I.M. performed analyses of Pluto's shape and gravity field and performed tectonics calculations. S.K. provided Love numbers for Pluto. J.K.S. provided estimates of volatile deposition timescales for Sputnik Planitia and provided input on volatile transport.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.T.K. (jkeane@lpl.arizona.edu).

METHODS

Could Sputnik Planitia be at the tidal axis by chance? To assess whether the alignment of Sputnik Planitia and the tidal axis is due to chance, we generated a large number ($N > 100,000$) of points on the surface of a unit sphere. The latitude, θ , and longitude, φ , of a randomly positioned point on a sphere can be written as $\theta = \cos^{-1}(2u - 1)$ and $\varphi = 2\pi v$, where u and v are random numbers spanning 0 and 1. We calculated the great-circle distance between each individual random point and the three nearest surface expressions of the three principal axes of inertia. Extended Data Fig. 7 shows the cumulative probability of a point being a certain angular distance away from one particular principal axis (for example, the tidal axis) in blue. The cumulative probability of a point being near one of two particular principal axes (for example, the tidal or spin axes) is orange. The cumulative probability of a point being near any one of the three principal axes is yellow. A simpler geometric method for calculating the probability of Sputnik Planitia being within angular distance, γ , of any single principal axis (blue line in Extended Data Fig. 7) involves taking the ratio of the area of two spherical caps with radius γ to the total surface area of the sphere. With this method, the probability is simply $1 - \cos(\gamma)$.

We determined the centre of Sputnik Planitia by fitting it to a small circle using the available global map of Pluto (Fig. 1). Using this technique, we found a centre of 176° E, 24° N, which is around 24° away from the tidal axis. There is a probability of about 9% that Sputnik Planitia is this close to the tidal axis (either at the anti-Charon point or at the sub-Charon point), and an approximately 26% chance that Sputnik Planitia would be close to any one of the principal axes. The authors of ref. 16 fitted Sputnik Planitia to an ellipse, and found a centre closer to the tidal axis, resulting in correspondingly smaller probabilities.

Predicted shape and gravity field of Pluto. We follow the notation of ref. 8, and describe Pluto's gravity field and shape in terms of spherical harmonic coefficients. However, an important correction must be made when investigating the tidal deformation of Pluto due to Charon. It is normally assumed that the perturbing body has a mass that is much larger than that of the deforming body (for example, Jupiter is much more massive than Europa). However, if the mass of the perturbing body is comparable to, or smaller than, the mass of the larger body (as is the case for Charon acting on Pluto), many of the gravity and shape equations must be augmented.

Ignoring TPW, the hydrostatic, non-zero degree-2 spherical harmonic gravity coefficients, $C_{2,0}$ (also commonly written as J_2) and $C_{2,2}$, can be written as:

$$J_2^{\text{HYD}} = -C_{2,0}^{\text{HYD}} = \frac{1}{6} k_2^{\text{T}'} q \left(\frac{2M + 5M_C}{M + M_C} \right)$$

$$C_{2,2}^{\text{HYD}} = \frac{1}{4} k_2^{\text{T}'} q \left(\frac{M_C}{M + M_C} \right)$$

where q is defined as $q \equiv (\Omega^2 R^3)/(GM)$, Ω is the rotation rate of Pluto/Charon, M is the mass of Pluto, M_C is the mass of Charon, G is the gravitational constant and $k_2^{\text{T}'}$ is the degree-2 long-term tidal Love number for the case without an elastic lithosphere. Love numbers (potential Love number k_2 and displacement Love numbers h_2 and ℓ_2) describe the response of a planet to the perturbing potential. Ignoring any contribution from TPW:

$$\frac{J_2^{\text{HYD}}}{C_{2,2}^{\text{HYD}}} = \frac{2}{3} \left(\frac{2M + 5M_C}{M_C} \right)$$

Note that this ratio is independent of the interior structure, and if $M_C \gg M$ (as is commonly assumed for tidally deformed planetary satellites) then $J_2^{\text{HYD}}/C_{2,2}^{\text{HYD}} = 10/3 = 3.3$. For Pluto, $J_2^{\text{HYD}}/C_{2,2}^{\text{HYD}} = 14.3$.

Ignoring TPW, the radial displacement due to hydrostatic rotational and tidal deformation can be written:

$$d_r(\theta, \varphi) = h_2^{\text{T}'} R q \left[-\frac{1}{3} P_{2,0}(\cos\theta) - \frac{1}{2} \left(\frac{M_C}{M + M_C} \right) P_{2,0}(\cos\theta) \right. \\ \left. + \frac{1}{4} \left(\frac{M_C}{M + M_C} \right) P_{2,2}(\cos\theta) \cos(2\varphi) \right]$$

where R is the radius of Pluto, $h_2^{\text{T}'}$ is the degree-2 long-term tidal displacement Love number for the case without an elastic lithosphere and $P_{2,0}$ is the degree-2 order-0 unnormalized associated Legendre function²⁶. The corresponding oblateness for Pluto is:

$$\frac{d_r(\theta = 90^\circ, \varphi = 0^\circ) - d_r(\theta = 0^\circ, \varphi = 0^\circ)}{R + d_r(\theta = 90^\circ, \varphi = 0^\circ)} \approx \frac{d_r(\theta = 90^\circ, \varphi = 0^\circ) - d_r(\theta = 0^\circ, \varphi = 0^\circ)}{R} \\ = 4.13 \times 10^{-4} \left(\frac{h_2^{\text{T}'}}{5/2} \right)$$

For our nominal Pluto interior structure model, $h_2^{\text{T}'} = 1.773$, and so the oblateness is 2.93×10^{-4} . The maximum possible oblateness would occur if Pluto were a uniform, hydrostatic fluid ($h_2^{\text{T}'} = 5/2$), corresponding to an oblateness of 4.13×10^{-4} . The New Horizons upper limit for Pluto's oblateness is an order of magnitude larger than either of these values¹³.

The inertia tensor of Pluto and TPW. TPW solutions can be found by diagonalizing the non-equilibrium inertia tensor. We follow ref. 8 to compute this inertia tensor, with the aforementioned modifications to account for the small M_C with respect to M . Ignoring the spherically symmetric contributions (which do not control orientation of a planet), the non-equilibrium inertia tensor can be written as:

$$I_{ij} = \left[\left(1 + k_2^{\text{L}} \right) M R^2 C_{2,0}^{\text{SP}*} \left(\frac{1}{3} \delta_{ij} - \hat{e}_i^{\text{SP}} \hat{e}_j^{\text{SP}} \right) \right] + \left[\left(k_2^{\text{T}'} - k_2^{\text{T}} \right) \frac{\Omega^2 R^5}{3G} \left(\hat{e}_i^{\text{R}'} \hat{e}_j^{\text{R}'} - \frac{1}{3} \delta_{ij} \right) \right] \\ + \left[\left(k_2^{\text{T}'} - k_2^{\text{T}} \right) \left(\frac{M_C}{M + M_C} \right) \frac{\Omega^2 R^5}{G} \left(\frac{1}{3} \delta_{ij} - \hat{e}_i^{\text{T}'} \hat{e}_j^{\text{T}'} \right) \right]$$

where the first bracketed term is the contribution from loading in Sputnik Planitia, the second bracketed term is the rotational deformation and the final term is the tidal deformation. Together, the last two bracketed terms represent the rotational and tidal components of the remnant figure; δ_{ij} is the Kronecker delta function, and \hat{e}^{SP} , $\hat{e}^{\text{R}'}$ and $\hat{e}^{\text{T}'}$ are unit vectors describing the centre of Sputnik Planitia, the initial rotation pole and the initial tidal pole (sub-Charon point), respectively. For example, if the spherical coordinates of Sputnik Planitia are $(\theta_{\text{SP}}, \varphi_{\text{SP}})$, then $\hat{e}^{\text{SP}} = (\sin\theta_{\text{SP}} \cos\varphi_{\text{SP}}, \sin\theta_{\text{SP}} \sin\varphi_{\text{SP}}, \cos\theta_{\text{SP}})$. $k_2^{\text{T}'}$ is the degree-2 loading Love number. All Love numbers here are the long-term Love numbers, or so-called fluid Love numbers. $C_{2,0}^{\text{SP}*}$ is the unnormalized degree-2 gravity coefficient of Sputnik Planitia for the case when it is centred at the north pole. Assuming an axisymmetric load with surface density σ and angular radius ψ :

$$C_{2,0}^{\text{SP}*} = \frac{2\pi R^2 \sigma}{M} \int_0^\psi d\theta \sin\theta P_{2,0}(\cos\theta)$$

where $J_2 \equiv -C_{2,0}$. Because the remnant figure provides stabilization against TPW, it is useful to define a normalized load size:

$$Q' = -\frac{C_{2,0}^{\text{SP}*}}{C_{2,0}^{\text{RF}}} = \frac{(1 + k_2^{\text{L}})}{(k_2^{\text{T}'} - k_2^{\text{T}})} \frac{GM}{\Omega^2 R^3} \frac{6(M + M_C)}{(2M + 5M_C)} C_{2,0}^{\text{SP}*}$$

where $C_{2,0}^{\text{RF}}$ is the unnormalized degree-2 zonal gravity coefficient associated with Pluto's remnant figure, which can be derived from converting the inertia tensor of the remnant figure into spherical gravity coefficients²⁷. Note that this normalization is different from the one used in refs 8 and 16 because it includes both the rotational and tidal deformation contributions to the remnant figure (hence our use of Q' instead of Q). Q' is related to Q by:

$$Q = \left(1 + \frac{3}{2} \frac{M_C}{M + M_C} \right) Q'$$

For the case of Pluto and Charon, the factor in front of Q' is 1.16. With this normalization, TPW solutions can be found by diagonalizing the normalized, non-equilibrium inertia tensor:

$$\bar{I}_{ij} = Q' \left(\frac{1}{3} \delta_{ij} - \hat{e}_i^{\text{SP}} \hat{e}_j^{\text{SP}} \right) + \frac{2(M + M_C)}{(2M + 5M_C)} \left(\hat{e}_i^{\text{R}'} \hat{e}_j^{\text{R}'} - \frac{1}{3} \delta_{ij} \right) \\ + \frac{6M_C}{(2M + 5M_C)} \left(\frac{1}{3} \delta_{ij} - \hat{e}_i^{\text{T}'} \hat{e}_j^{\text{T}'} \right)$$

The interior structure dependence is described by the Love numbers factor $(1 + k_2^{\text{L}})/(k_2^{\text{T}'} - k_2^{\text{T}})$ in the definition of Q' . Both the numerator and denominator of this factor are sensitive to the interior structure, but this dependence almost disappears when the ratio is used. This is shown clearly in Extended Data Fig. 5k, and further enhances the robustness of our solutions.

We compute the long-term Love numbers by solving the mass, momentum and Poisson equations for the deformation of a spherically symmetric, non-rotating, elastic and isotropic body^{14,28,29}. As discussed in the main text, we considered a four-layer interior structure model, consisting of a silicate-rich solid core (density: 3.36 g cm^{-3} ; radius: 858 km), liquid water ocean (density: 1.0 g cm^{-3} ; thickness: 10 km), overlaid by a two-layer water-ice mantle (density: 0.95 g cm^{-3} ; thickness: 320 km). The ice mantle is comprised of a weak, viscous lower mantle, beneath an elastic lithosphere of varying elastic thickness (0–70 km), rigidity (3.49 GPa) and bulk modulus (9.3 GPa) (ref. 30). The long-term Love numbers can be computed

by considering the infinite time limit. Alternatively, these Love numbers can be computed by assuming that all regions, with the exception of the elastic lithosphere, are inviscid. For simplicity, we assume infinite viscosity in all layers, zero frequency dependence and that only the upper mantle is elastic (all common assumptions). Love numbers decrease as the elastic thickness or rigidity of the lithosphere decrease. For our nominal interior structure with elastic thicknesses spanning 0 to 70 km, k_2^T spans 0.7653 to 0.4682; k_2^L spans -0.9856 to -0.4500 . Similarly, the tidal displacement Love numbers, h_2 and ℓ_2 , can be evaluated. Over the same span in elastic thickness, h_2^T spans 1.7509 to 0.9182; h_2^L spans -3.2230 to -1.7345 ; ℓ_2^T spans 0.4374 to 0.9182; and ℓ_2^L spans -0.7847 to -0.3858 . Our Love numbers differ slightly from those of ref. 16, primarily owing to different assumptions about the core structure. Our results are not strongly sensitive to assumed Love numbers and interior structure.

Tectonics patterns due to global expansion, TPW and loading of Sputnik Planitia. Following ref. 30, the stresses on the surface of a planet due to a generalized displacement $\mathbf{d} = (d_r, d_\theta, d_\varphi)$ in spherical coordinates can be written:

$$\begin{aligned}\tau_{\theta\theta} &= \frac{\lambda}{\lambda + 2\mu} \tau_{rr} + \frac{2\mu}{R(\lambda + 2\mu)} \left[(3\lambda + 2\mu)d_r + 2(\lambda + \mu)\partial_\theta d_\theta \right. \\ &\quad \left. + \lambda \left(\frac{\partial_\varphi d_\varphi}{\sin\theta} + d_\theta \cot\theta \right) \right] \\ \tau_{\varphi\varphi} &= \frac{\lambda}{\lambda + 2\mu} \tau_{rr} + \frac{2\mu}{R(\lambda + 2\mu)} \left[(3\lambda + 2\mu)d_r + \lambda\partial_\theta d_\theta \right. \\ &\quad \left. + 2(\lambda + \mu) \left(\frac{\partial_\varphi d_\varphi}{\sin\theta} + d_\theta \cot\theta \right) \right] \\ \tau_{\theta\varphi} &= \frac{\mu}{R} \left[\partial_\theta d_\varphi + \frac{\partial_\varphi d_\theta}{\sin\theta} - d_\varphi \cot\theta \right]\end{aligned}$$

where we assume a compressible interior, and μ and λ are the Lamé parameters. Note that $\tau_{rr} = 0$ for displacements due to rotational or tidal deformation.

From the previous equations, the isotropic stress due to an isotropic radius change, δR , can be written as:

$$\tau_{\theta\theta} = \tau_{\varphi\varphi} = 2\mu \frac{(2\mu + 3\lambda)}{(\lambda + 2\mu)} \frac{\delta R}{R} = \frac{E}{(1 - \nu)} \frac{\delta R}{R}$$

where E and ν are the elastic modulus and Poisson's ratio, respectively.

The displacement due to rotational and tidal deformation associated with TPW can be written as:

$$(d_r, d_\theta, d_\varphi) = R \sum_{m=0}^l \sum_{i=1}^2 \left(h_2, \ell_2 \partial_\theta, \frac{\ell_2 \partial_\varphi}{\sin\theta} \right) \left(U_{2,m,i}^f - U_{2,m,i}^i \right) Y_{l,m,i}$$

where U^i and U^f are the initial (pre-TPW) and final (post-TPW) gravitational potentials, respectively. $Y_{l,m,i}$ are the spherical surface harmonics:

$$\begin{pmatrix} Y_{l,m,1}(\theta, \varphi) \\ Y_{l,m,2}(\theta, \varphi) \end{pmatrix} = P_{l,m}(\cos\theta) \begin{pmatrix} \cos m\varphi \\ \sin m\varphi \end{pmatrix}$$

where l and m are the spherical harmonic degree and order, respectively. The gravitational potential expansion coefficients are:

$$\begin{aligned}U_{2,m,i}^f &= (2 - \delta_{m0}) \frac{(2-m)!}{(2+m)!} \left[-\frac{\Omega_i^2 R^3}{3GM} Y_{2,m,i}(\theta_R^f, \varphi_R^f) \right. \\ &\quad \left. + \frac{M_C}{M + M_C} \frac{\Omega_i^2 R^3}{GM} Y_{2,m,i}(\theta_T^f, \varphi_T^f) \right] \\ U_{2,m,i}^i &= (2 - \delta_{m0}) \frac{(2-m)!}{(2+m)!} \left[-\frac{\Omega_i^2 R^3}{3GM} Y_{2,m,i}(\theta_R^i, \varphi_R^i) \right. \\ &\quad \left. + \frac{M_C}{M + M_C} \frac{\Omega_i^2 R^3}{GM} Y_{2,m,i}(\theta_T^i, \varphi_T^i) \right]\end{aligned}$$

where $(\theta_R^f, \varphi_R^f)$ and $(\theta_T^f, \varphi_T^f)$ are the spherical coordinates of the final rotation pole and the tidal pole (sub-Charon point), respectively. $(\theta_R^i, \varphi_R^i)$ and $(\theta_T^i, \varphi_T^i)$ are the spherical coordinates of the initial rotation pole and tidal pole (sub-Charon point). Ω_i^i and Ω_i^f are the initial and final rotation rates.

The displacement due to mass loading in Sputnik Planitia can be written as:

$$(d_r, d_\theta, d_\varphi) = R \sum_{l=0}^{\infty} \sum_{m=0}^l \sum_{i=1}^2 \left(h_l^L, \ell_l^L \partial_\theta, \frac{\ell_l^L \partial_\varphi}{\sin\theta} \right) U_{l,m,i}$$

Assuming an axisymmetric load, the gravitational potential expansion coefficients are:

$$U_{l,m,i} = (2 - \delta_{m0}) \frac{(2-m)!}{(2+m)!} \frac{2\pi R^2 \sigma}{M} Y_{l,m,i}(\theta_{SP}, \varphi_{SP}) \int_0^\psi d\theta \sin\theta P_{l,0}(\cos\theta)$$

and the radial stress due to mass loading in Sputnik Planitia can be written as:

$$\tau_{rr} = -\frac{gM}{4\pi R^2} \sum_{l=0}^{\infty} \sum_{m=0}^l \sum_{i=1}^2 (2l+1) U_{l,m,i} Y_{l,m,i}(\theta_{SP}, \varphi_{SP})$$

where g is the surface gravity of Pluto ($g = GM/R^2$).

Tectonics patterns are derived from the eigenvalue of the local stress tensor³¹, which can be constructed by summing all stress components mentioned above. Normal faulting is perpendicular to the maximum principal stress.

Rotational stability of Pluto. If volatiles tend to accumulate at Pluto's north and south poles, this may make Pluto more susceptible to polar wander than previously expected²¹. As a polar ice cap grows in mass, it effectively reduces the dynamical flattening of the planet. Once the polar caps exceed the tidal rotational flattening, the maximum and intermediate principal axes of inertia will swap—reorienting Pluto by 90° around the tidal axis, placing the former polar cap on the intermediate principal axis of inertia (the orbital axis at (90° E, 0° N) or (270° E, 0° N) as in Fig. 1). If we assume that volatiles are sequestered into two 60°-radius polar caps, with ice densities of 1 g cm^{-3} , then this rotational stability limit occurs when the ice caps reach a thickness of around 180 m (equivalent to a global layer of around 90 m). This calculation assumes that the polar ice caps can be partially compensated by lithospheric flexure. If caps are not partially supported (or if they are constructed more rapidly than the lithosphere can adjust), then the threshold is smaller: about 80 m (equivalent to a global layer of about 40 m). Note that the total volume of ice required within these 'destabilizing' ice caps is comparable to the expected volume of volatile ice within Sputnik Planitia.

If polar ice caps can grow large enough to destabilize Pluto, they would result in substantially different tectonic patterns—perhaps akin to 'crop-circle' faults on Europa, which are hypothesized to result from a similar style of reorientation³² (albeit due to differential crustal thickening, and not surface volatile transport). The lack of such an observed tectonic pattern may suggest that this style of reorientation has not occurred on Pluto. Perhaps the volatile reservoir on Pluto is too small to build such a polar cap, or the remnant figure is larger. Although Pluto's polar regions are indeed rich in volatiles³, the thickness of volatiles and their compensation state are unclear. Some eroded mantled terrains in the northern uplands have thicknesses of approximately 1 km (ref. 2). If these deposits ever covered large fractions of polar regions, they would easily destabilize Pluto. If polar volatiles are sufficiently mobile, then it is conceivable that they may adjust to changes in the planet's insolation conditions faster than the planet can reorient. In such a case, polar volatiles would act more similarly to a hydrostatic tidal rotational bulge, and may not be capable of controlling the planet's orientation over long timescales.

Sputnik Planitia as a cold trap. As Sputnik Planitia is a large topographic low, and there is no observed troposphere on Pluto³³ (that is, temperature decreases with altitude on Pluto), we posit that Sputnik Planitia is an intrinsic cold trap. Other factors may contribute to Sputnik Planitia being a cold trap, such as atmospheric circulation in response to its topography or intrinsic albedo variations associated with the underlying basin. This hypothesis is explored in detail in ref. 34. Over time, scattered volatile ice deposits should migrate into one large deposit (Sputnik Planitia) owing to lower-albedo terrains that surround the higher-albedo ice. The darker terrains have higher surface temperatures, resulting in an enhancement of volatile ice sublimation from the margins of the ice deposits. Thus, if ice condensation is constant across the interiors of multiple ice sheets, then smaller ice sheets experience a larger fractional loss of ice from their margins, resulting in their eventual disappearance and atmospheric transport of their volatile content to a single dominant ice deposit (a process we term 'oligarchic growth of cold traps').

If reorientation is controlled by the accumulation of volatiles within Sputnik Planitia, then the approximate reorientation timescale can be estimated from the time required to fill Sputnik Planitia with volatiles. From Fig. 2b and Extended Data Fig. 3, the typical volatile ice thicknesses required to drive reorientation are around 5 km. Following ref. 35, it would take approximately 5 million years to grow a 5 km N_2 ice cap given Pluto's present average atmospheric pressure and temperature³³.

Similar timescales can be estimated by considering sublimation mass fluxes due to seasonal differences in insolation⁷.

Seasonal wobbles of Pluto due to mass transport into and out of Sputnik Planitia. If volatiles migrate into and out of Sputnik Planitia on seasonal timescales, then one may expect corresponding small-scale oscillations in the planet's principal axes of inertia. These small-amplitude wobbles would be analogous to atmospheric-pressure-driven wobbles on Earth²⁵. Extended Data Fig. 8a, b shows how Pluto's minimum and maximum principal axes of inertia would change orientation in response to small changes in the thickness of volatile ices within Sputnik Planitia. If these changes occurred seasonally, then Pluto would not have time to damp the tidal rotational energy (TPW timescales are generally of the order of millions of years), and the principal axes would wobble about the tidal and rotational axes^{25,36}. In a reference frame fixed to Pluto, such wobbles would result in apparent motions of Pluto in the sky (even though Pluto and Charon would remain in almost perfect double synchronous spin-orbit lock). Such motions would result in tidal heating very similar to obliquity and/or libration tides, which can be important for the heat budget of icy satellites³⁷. Because these wobbles are driven by the continual sublimation and redeposition of volatiles on Pluto, they may be persistent, resulting in small but continual amounts of extra heating within Pluto.

The amount of heat generated from these postulated seasonal wobbles can be estimated by determining the change in rotational kinetic energy in response to the transport of ice into and out of Sputnik Planitia. Rotational kinetic energy, KE, is related to the planet's spin vector, \mathbf{w} , and its inertia tensor, \mathbf{I} : $KE = \frac{1}{2} \mathbf{w}^T \mathbf{I} \mathbf{w}$. For Pluto's current rotation rate (6.39 Earth days), and typical assumptions of Pluto's figure and moments of inertia (described above), Pluto's rotational kinetic energy is about 4.7×10^{23} J. Adding or removing one metre of ice (approximately the amount of ice transported seasonally across Pluto⁷) from Sputnik Planitia changes this rotational kinetic energy by approximately 10^{17} J (assuming the rotation rate remains unchanged) (Extended Data Fig. 8c). Assuming this occurs over one Pluto year (248 Earth years), this translates into a heat production rate of 5×10^7 J s⁻¹. This is approximately 1/1,000 of the present-day radiogenic heating expected within Pluto³⁸. Although the expected heat production rate is low, this heat may be released closer to the surface (for example, localized in crustal faults), whereas radiogenic heating is probably concentrated within Pluto's silicate core, deep beneath the surface. Future work will need to quantify the mechanics and efficacy of this 'wobble-heating' process in more detail.

The magnitude and geometry of any such wobbles depend on both the amount of material being transported and the differences between Pluto's moments of inertia (equivalently, its degree-2 spherical harmonic gravity coefficients²⁷). If the TPW hypothesis for Sputnik Planitia is correct, then Pluto's present-day degree-2 gravity has some contributions from Sputnik Planitia and Pluto's reoriented remnant figures. The exact values depend on the proposed reorientation event. Extended Data Fig. 8d–f shows predicted, present-day spherical harmonic gravity coefficients for the reorientation geometries summarized in Fig. 2a. Future measurement of degree-2 gravity of Pluto, be it from spacecraft or remote observations of wobbles, may further constrain which TPW models for Sputnik Planitia are viable.

Other geologic evidence for reorientation. Sublimation often creates geologic features that align north–south. Although several terrains on Pluto are characterized by elongated, plausibly sublimation-driven landforms (for example, pitted uplands and the bladed terrain of Tartarus Dorsa²), not all align north–south. The washboard terrain northwest of Sputnik Planitia is characterized by parallel ridges and troughs that trend southwest–northeast². In our pre-TPW view of Pluto, these features become aligned north–south (Extended Data Fig. 4a, b). Relative age-dating of these features and other sublimation structures may provide constraints on the timing of the formation of Sputnik Planitia.

If the reorientation of Pluto was due to the gradual infill of volatiles within Sputnik Planitia (as in Fig. 3d–i), then the stress and tectonic patterns may also change gradually with time. The time evolution of the tectonic patterns predicted by our simple reorientation model (Fig. 3d–i) are shown in Extended Data Fig. 6 and Supplementary Videos 1–3. A nearly ubiquitous result in these tectonic models is that at a certain distance away from Sputnik Planitia, faults transition from being quasi-radial to Sputnik Planitia to quasi-azimuthal. This transition marks a change in the dominant source of stress; loading stresses dominate near Sputnik Planitia, whereas reorientation stresses dominate far away. We use this transition as evidence of the reorientation to explain the sharp change in fault azimuths observed on Pluto (Fig. 2g–k). Curiously, the location of this transition changes as a function of the loading within Sputnik Planitia in these time-sequence models. In particular, the transition moves progressively further away from Sputnik Planitia as it is filled with volatiles (Extended Data Fig. 6a–f). Eventually, for very thick deposits within Sputnik Planitia, the loading stresses dominate reorientation stresses globally (Extended Data Fig. 6g, h). This change in stress field with time may be recorded in the cross-cutting relationships of faults on Pluto. For example,

we may predict that within the faults far from Sputnik Planitia (Djaggawul and the 'spider'), faults closer to Sputnik Planitia will be older, and may be cross-cut by faults radial to Sputnik Planitia. Further geologic mapping of Pluto (even in regions far from Sputnik Planitia), may provide valuable insight into the nature and origin of Sputnik Planitia.

The mass anomaly and underlying structure of Sputnik Planitia. As noted in the main text, the mass anomaly from a 3-km-deep, uncompensated topographic depression has a mass anomaly of $Q' = -4.14$ (Fig. 2b), which is too large and negative for Sputnik Planitia to be at its present location. Sputnik Planitia must have a mass anomaly between -0.3 and 1.8 to be at its present location. With $Q' = -4.14$, Sputnik Planitia would reorient almost completely to the north pole. Adding volatiles into the basin (while maintaining a 3-km-deep basin) reduces the mass anomaly of the basin. However, for typical values for the density of the Pluto's ice-rich crust (930 kg m^{-3}) and the predominantly N₂ volatile ice within the basin ($1,000 \text{ kg m}^{-3}$), the thickness required to give $-0.3 < Q' < 1.8$ is well above the estimated thickness of the volatile ice ($< 10 \text{ km}$; refs 2, 4, 5). Although decreasing the density of the crust and increasing the density of the volatile infill can reduce the required thickness (Extended Data Fig. 3b), it is difficult (it would require a very dense volatile infill and extremely low-density crust). Thus, for Sputnik Planitia to be at its present location, the underlying basin must at least be partially compensated.

If Sputnik Planitia is an impact basin, then there are several processes that could plausibly compensate for the mass anomaly associated with the topographic depression. These processes can include mantle uplift, isostatic adjustment of the basin and surrounding terrains, cooling and contraction of impact melt, changes in porosity and the emplacement of ejecta¹⁵. On the Moon and other terrestrial planets, these processes often give impact basins stochastically positive and negative total mass anomalies, despite being topographic lows (Extended Data Fig. 1; ref. 12). Even if impact processes on icy planets and satellites are somehow fundamentally different than those on their terrestrial counterparts, unavoidable components such as ejecta blankets can considerably offset the negative mass anomaly from the topographic depression. This may be why large impact basins on other icy bodies do not always result in poleward reorientation (for example, Herschel on Mimas¹⁷). As an example, the reorientation of the Moon due to the formation of the South Pole–Aitken impact basin was more strongly controlled not by the basin itself, but by the deposition of a thick ejecta blanket on the far side of the Moon^{12,39}.

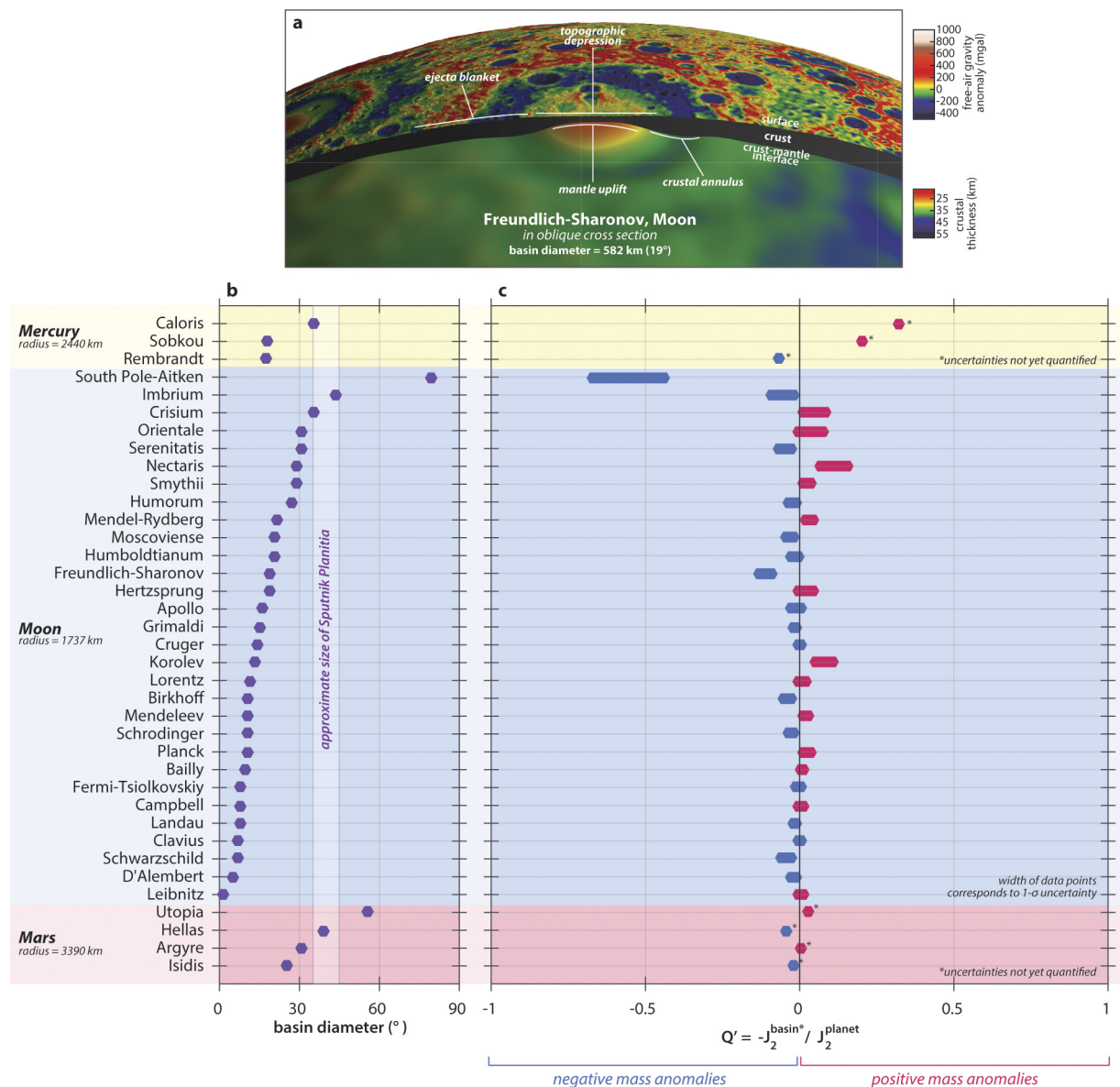
Extended Data Fig. 3 showcases four possible simple structures for the Sputnik Planitia basin, and their impact on its total mass anomaly. The first model (Extended Data Fig. 3a) looks at volatiles loading an underlying basin that has an intrinsic mass anomaly of $Q' = 0$, which is dynamically equivalent to volatiles loading the surface of the planet. Since the Q' value of mass anomalies on other planets is seemingly random (Extended Data Fig. 1; ref. 12), this model serves as our null hypothesis, and is what we use to estimate ice thicknesses for loading and tectonic calculations. The second model (Extended Data Fig. 3b) considers ice filling a basin of fixed depth, and is already discussed above. The third model (Extended Data Fig. 3c) is a simple impact basin model (based on ref. 40) consisting of ice filling a basin of fixed depth, surrounded by an ejecta blanket with a total volume set by the total amount of material excavated from within the basin. The final model (Extended Data Fig. 4d) consists of a basin that is initially isostatically compensated by an uplift in a subsurface ocean, and subsequently filled with volatiles. This final model is examined more thoroughly in refs 16 and 41. Additional models can be constructed by summing components from the mass anomaly menu in Extended Data Fig. 2.

The inclusion of an ejecta blanket substantially offsets the mass anomaly associated with the topographic low of the Sputnik Planitia basin—from $Q' \approx -4$ to $Q' \approx -2$ (Fig. 2b, Extended Data Fig. 3c). This is still too negative for Sputnik Planitia to be at its present location (requiring $0.3 < Q' < 1.8$). Thus, like the model without ejecta (Extended Data Fig. 3b), this then requires that Sputnik Planitia must have some extra positive mass anomaly contributing to its total mass anomaly. However, including impact ejecta enables the volatiles within the basin to play a much larger role in the total mass anomaly for three reasons. First, adding an ejecta blanket changes the depth of the basin with respect to the mean radius of Pluto; we assume that the depth of the basin is measured with respect to the crater rim. Second, by reducing the mass anomaly of the underlying basin, it reduces the amount of volatiles required within the basin to reach an acceptable Q' value. Lastly, as the thickness of the volatiles within the basin increases, the total excavated volume of Sputnik Planitia increases, thus increasing the thickness of ejecta blanket. This results in steeper curves for Q' as a function of volatile ice thickness in Extended Data Fig. 2c compared with Extended Data Fig. 2b. Taking the ejecta blanket into account reduces the required volatile ice thickness from tens or hundreds of kilometres from the simple basin model (Extended Data Fig. 2b) to less than 10 km, which is consistent with other estimates for the thickness of volatiles within Sputnik Planitia^{2,4,5}. If the thickness of volatiles

within Sputnik Planitia is larger, then Sputnik Planitia must have an additional negative mass anomaly to enable it to be at its present location.

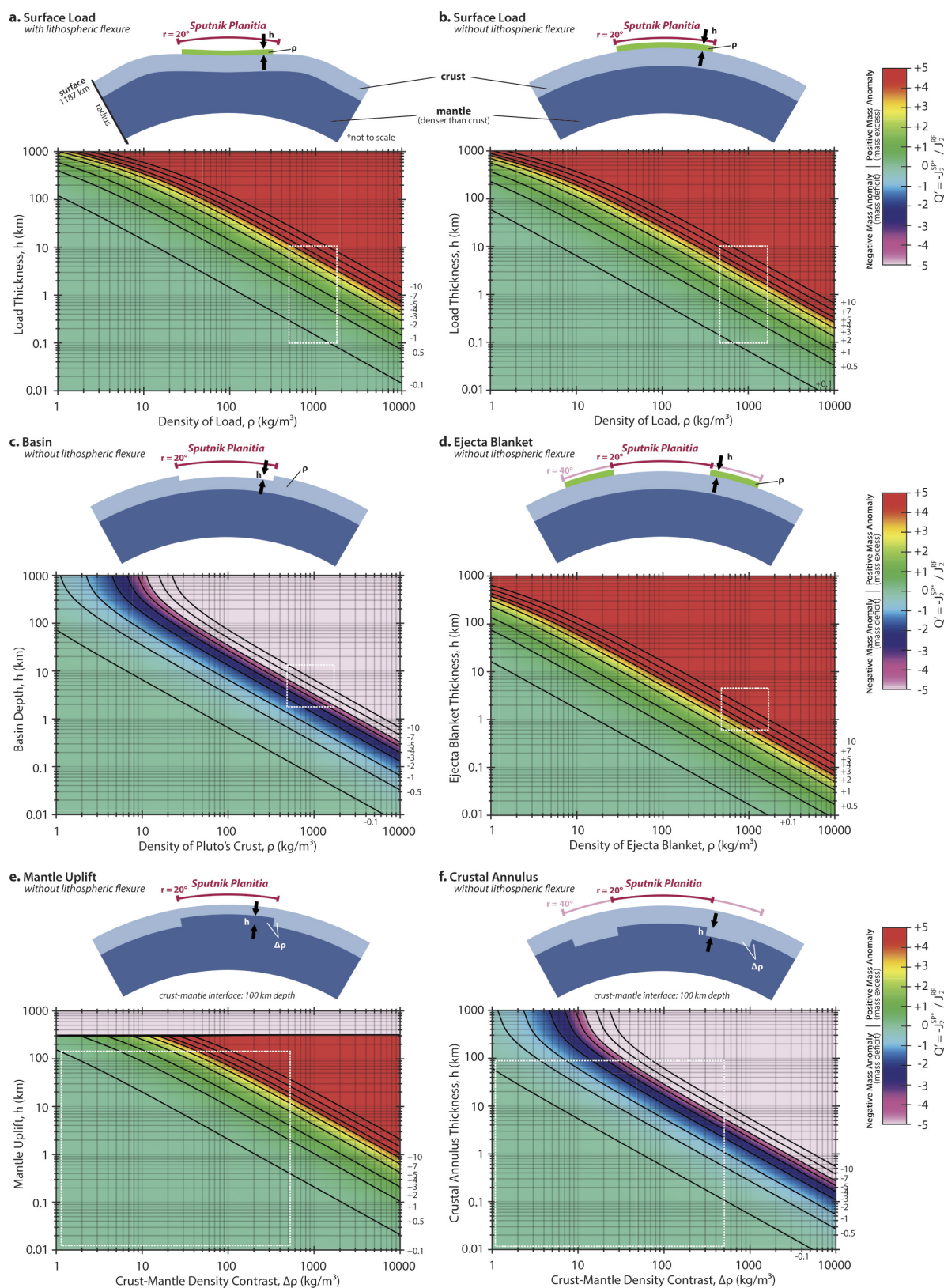
The inclusion of an uplift in the postulated subsurface ocean results in a similar Q' value to models with an ejecta blanket. If the initial (pre-volatile filled) Sputnik Planitia basin is isostatically compensated (via Airy isostasy), then the initial mass anomaly of the basin reduces from $Q' \approx -4$ to $Q' \approx -2$ (Fig. 2b, Extended Data Fig. 3d). It is important to note that despite the fact that this initial basin is 'compensated' it does not have $Q' = 0$. This arises from how a compensated structure fundamentally alters the inertia tensor (and thus Q'). Because the structure compensating for the topographic low is deeper within the planet, it contributes less to the inertia tensor (inertia tensors scale with radial distance squared), and thus the negative mass anomaly arising from the topographic low still dominates. Nonetheless, this reduction of the underlying basin's Q' value is sufficient to allow volatiles to overcome the remaining negative mass anomaly for ice thicknesses of 1–10 km. This mechanism is discussed much more rigorously in ref. 16. Continued study of New Horizons data, as well as thorough impact simulation studies, is needed to truly disentangle the possible cause of the positive mass anomaly within Sputnik Planitia.

26. Arfken, G. B. & Weber, H. J. *Mathematical Methods for Physicists* 7th edn 797–798 (Academic, 1995).
27. Lambeck, K. *The Earth's Variable Rotation: Geophysical Causes and Consequences* (Cambridge Univ. Press, 1980).
28. Takeuchi, H. & Saito, M. in *Methods in Computational Physics* (ed. Bolt, B. A.) 217–295 (Academic Press, 1972).
29. Sabadini, R., Vermeersen, B. & Cambiotti, G. *Global Dynamics of the Earth: Applications of Viscoelastic Relaxation Theory to Solid-Earth and Planetary Geophysics* (Springer, 2016).
30. Wahr, J. *et al.* Modeling stresses on satellites due to nonsynchronous rotation and orbital eccentricity using gravitational potential theory. *Icarus* **200**, 188–206 (2009).
31. Anderson, E. M. *The Dynamics of Faulting* (Oliver & Boyd, 1951).
32. Schenk, P., Matusyama, I. & Nimmo, F. True polar wander on Europa from global-scale small-circle depressions. *Nature* **453**, 368–371 (2008).
33. Gladstone, G. R. The atmosphere of Pluto as observed by New Horizons. *Science* **351**, aad8866 (2016).
34. Bertrand, T. & Forget, F. Observed glacier and volatile distribution on Pluto from atmosphere–topography processes. *Nature* <http://dx.doi.org/10.1038/nature19337> (2016).
35. Langmuir, I. The vapor pressure of metallic tungsten. *Phys. Rev.* **2**, 329–342 (1913).
36. Spada, G., Sabadini, R. & Boschi, E. Long-term rotation and mantle dynamics of the Earth, Mars, and Venus. *J. Geophys. Res.* **101**, 2253–2266 (1996).
37. Nimmo, F. & Spencer, J. R. Powering Triton's recent geologic activity by obliquity tides: implications for Pluto geology. *Icarus* **246**, 2–10 (2015).
38. McKinnon, W. B., Simonelli, D. P. & Schubert, G. in *Pluto and Charon* (eds Stern, S. A. & Tholen, D. J.) 295–346 (Univ. Arizona Press, 1997).
39. Kendall, J. D., Johnson, B. C., Bowling, T. J. & Melosh, H. J. Ejecta from south pole-Aitken basin-forming impact: dominant source of farside lunar highlands. In *46th Lunar Planet. Sci. Conf.* 2765 (Lunar and Planetary Institute, 2015).
40. Melosh, H. J. Large impact craters and the Moon's orientation. *Earth Planet. Sci. Lett.* **26**, 353–360 (1975).
41. Johnson, B. C., Bowling, T. J., Trowbridge, A. J. & Freed, A. M. Formation of the Sputnik Planum basin and the thickness of Pluto's subsurface ocean. *Geophys. Res. Lett.* **43**, 10068–10077 (2016).



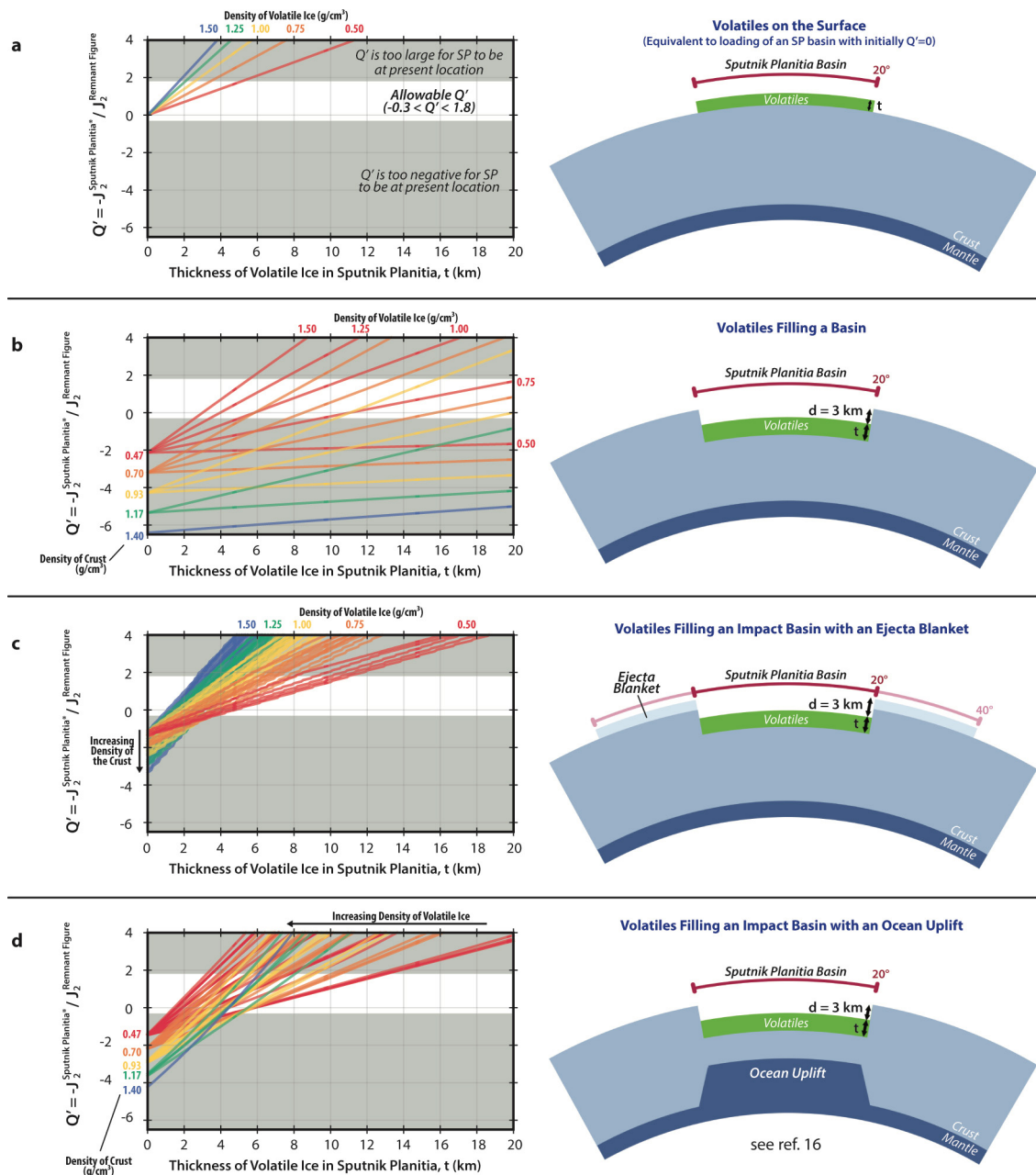
Extended Data Figure 1 | Mass anomalies of impact basins on other planetary bodies. a, An oblique cross-section of a typical lunar mascon basin, Freundlich–Sharonov, highlighting the different components that contribute to a basin’s overall mass anomaly. **b,** A table of large basins on other planetary bodies (in order of the basin’s diameter relative to the host planet) for which we have adequate gravity measurements to determine

their respective mass anomalies. **c,** The mass anomaly of each basin. The mass anomalies for lunar impact basins have been comprehensively characterized in ref. 12. Mass anomalies for impact basins on Mars and Mercury are calculated in the same way, using available gravity data for these two bodies. Uncertainties have not been quantified for impact basins on Mars and Mercury.



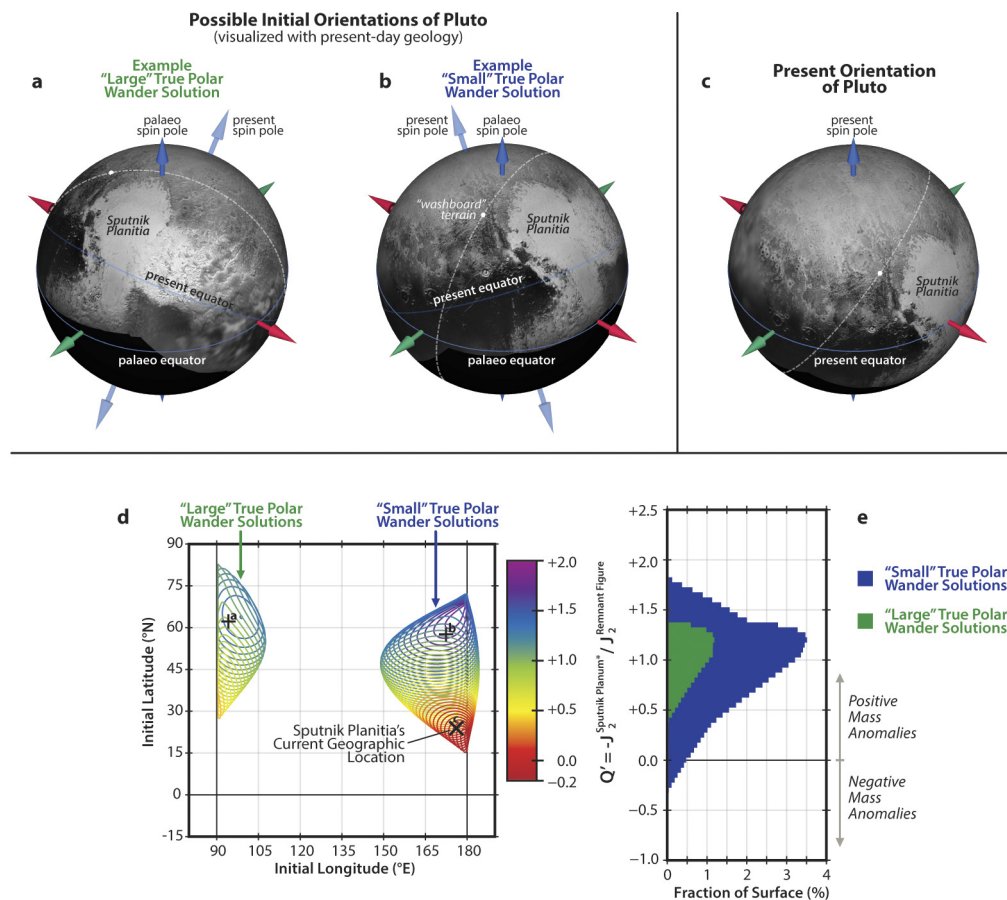
Extended Data Figure 2 | Mascon components and their mass anomalies. a–f, Mass anomaly (Q') of an individual component of a basin (for example, volatile ice load or topographic depression) as a function of that component's thickness and density. White dashed regions denote plausible areas of parameter space (densities of ices on Pluto are from

ref. 7; basin depths and ice thicknesses are from refs 1, 2, 4, 5; ejecta blanket thicknesses are estimated by redistributing the mass excavated from the basin into an annulus outside the basin, between 1 and 2 crater radii). The mass anomaly of the impact basin can be constructed by linearly summing these components.



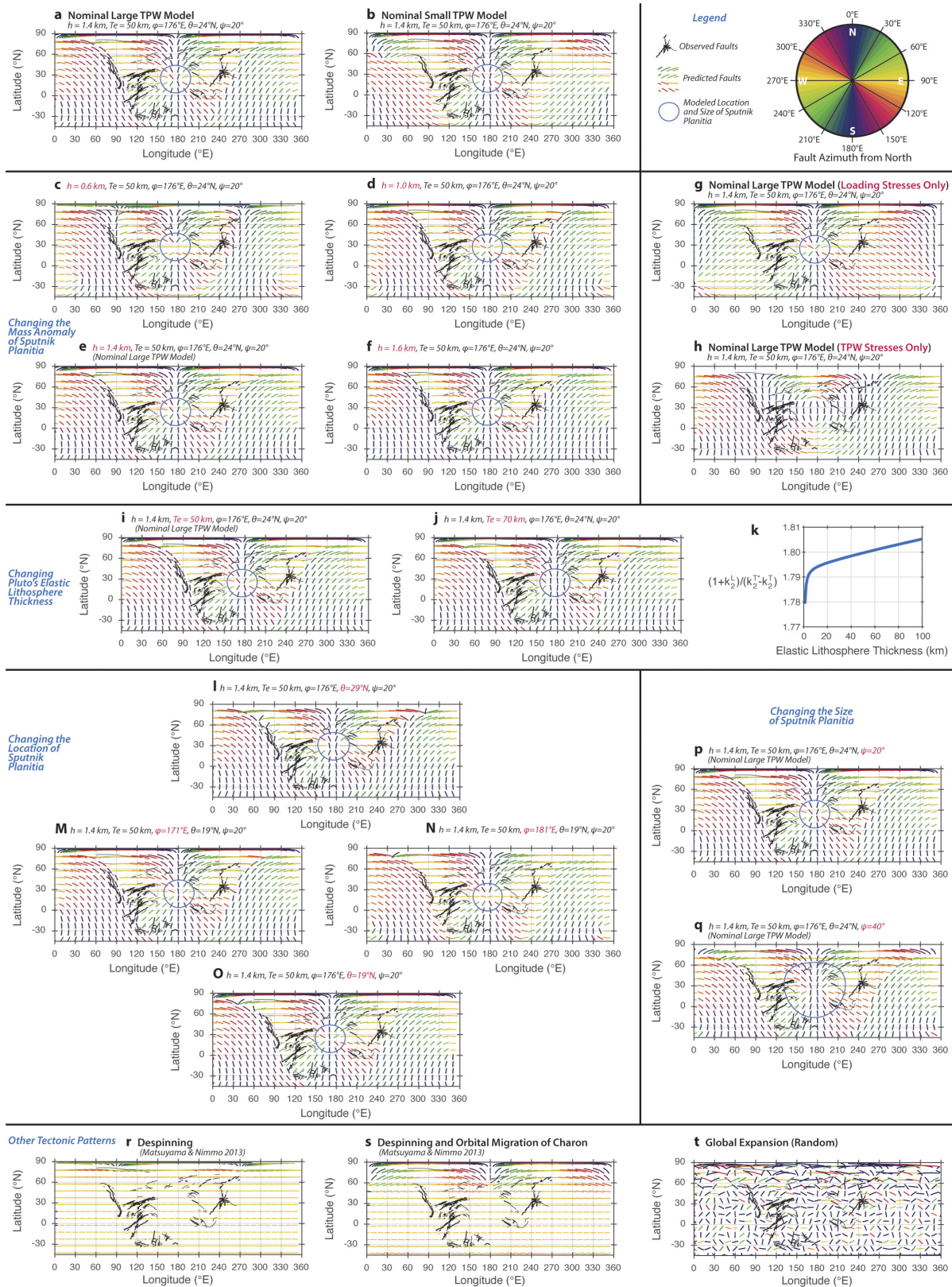
Extended Data Figure 3 | Simple models for Sputnik Planitia. **a–d**, Q' as a function of volatile thickness (left column) for each simple model of Sputnik Planitia (right column). Different colours and lines denote different model results assuming different densities of model components (volatile ice density, crust density, mantle density, and ejecta density, when appropriate). Nominal outputs from each model are shown in Fig. 2b. See Methods for discussion of this figure. **a**, Volatiles loading on the surface of a planet, which is equivalent to volatiles loading a basin with no intrinsic mass anomaly. **b**, Volatiles filling an initially uncompensated

basin, with a fixed depth of 3 km from the surface of the planet to the top of the volatiles. **c**, Volatiles filling an initially uncompensated impact basin surrounded by an ejecta blanket extending from 1 to 2 crater radii containing the total mass excavated from within the basin. The height from the top of the rim to the top of the volatiles is fixed to 3 km. **d**, Volatiles filling a basin that is initially compensated from isostatic uplift of the presumed ocean at depth (see ref. 16 for a more thorough investigation of this hypothesis). In all plots, it is assumed that volatiles are partially supported by Pluto's elastic lithosphere (Methods).



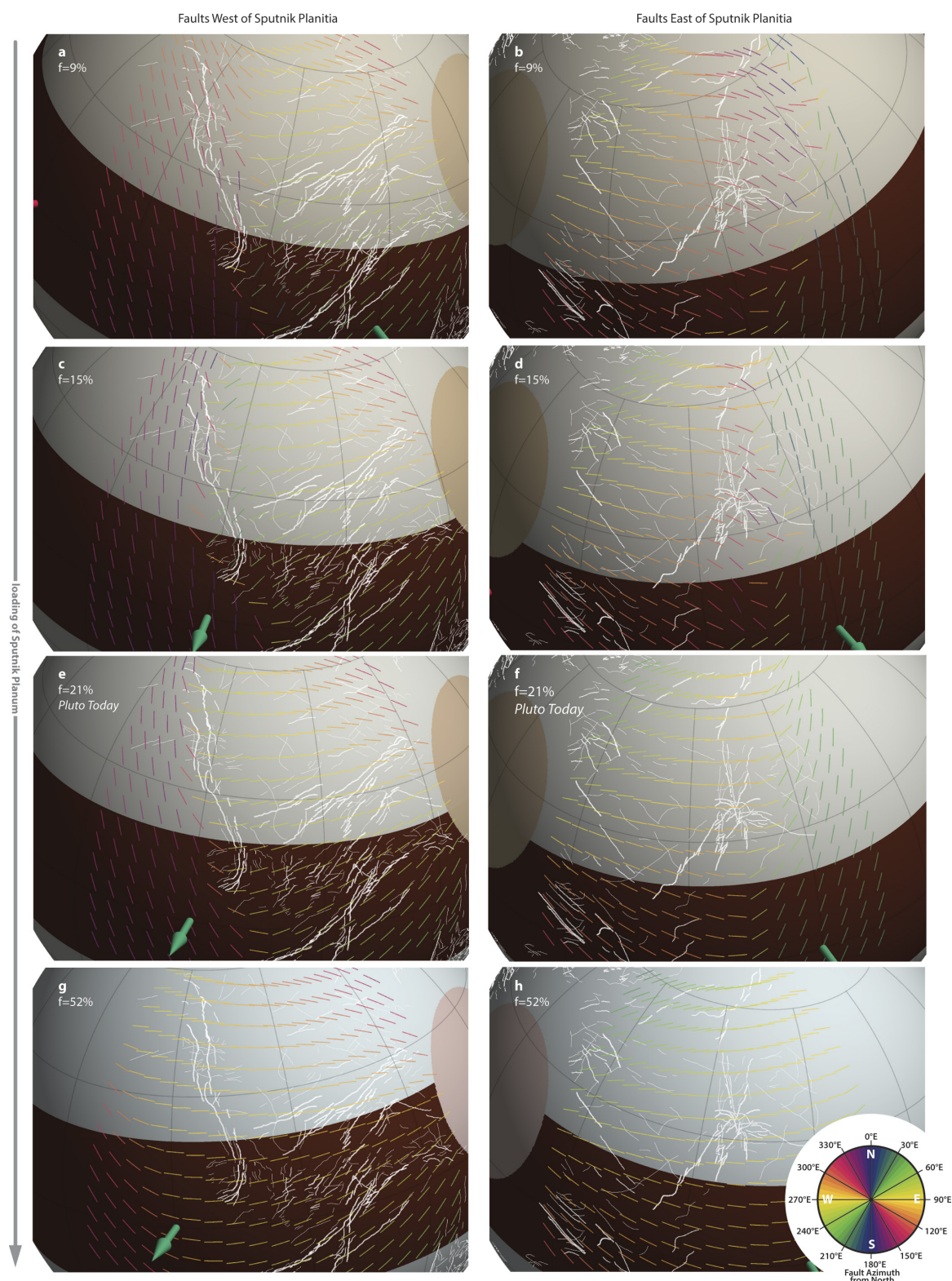
Extended Data Figure 4 | The initial orientation of Pluto and the range of possible mass anomalies. **a–c**, Orthographic spherical projections of Pluto for example initial orientations from the perspective of an inertial viewer, fixed with respect to the tidal/rotational axes. Base map: NASA/ Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute. The dashed white line indicates the strike of the 'washboard

terrain' (Methods). **d**, Contours enclosing the possible initial locations of Sputnik Planitia as a function Q' . This is the same as Fig. 2a, but in an equirectangular map projection. **e**, Histogram of the Q' values of allowable reorientation scenarios as shown in **d** (and Fig. 2a). The vast majority of solutions are positive mass anomaly solutions.



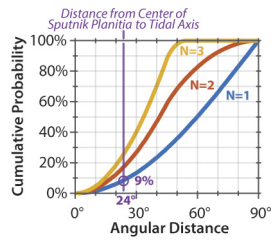
Extended Data Figure 5 | Tectonic models. a–q. Tectonic patterns depend on the geometry of proposed reorientation (a–f), the interior structure (i–k), and the size and location of the perturbing load (l–q). Yet, despite all of the possible dependencies, it is striking that for the allowed reorientation scenarios, the predicted tectonic patterns show little variance. Most predict quasi-radial faults proximal to Sputnik Planitia (due to loading, g),

transitioning to quasi-azimuthal faults distal to Sputnik Planitia (due to TPW stresses, h). Faults are coloured by azimuth, as in Fig. 2g–j. Black lines show mapped faults, as in Fig. 2g, h. r–t, Tectonics patterns from TPW, loading and global expansion provide a far better match to the observed fault distribution than do de-spinning, orbit migration or global expansion.

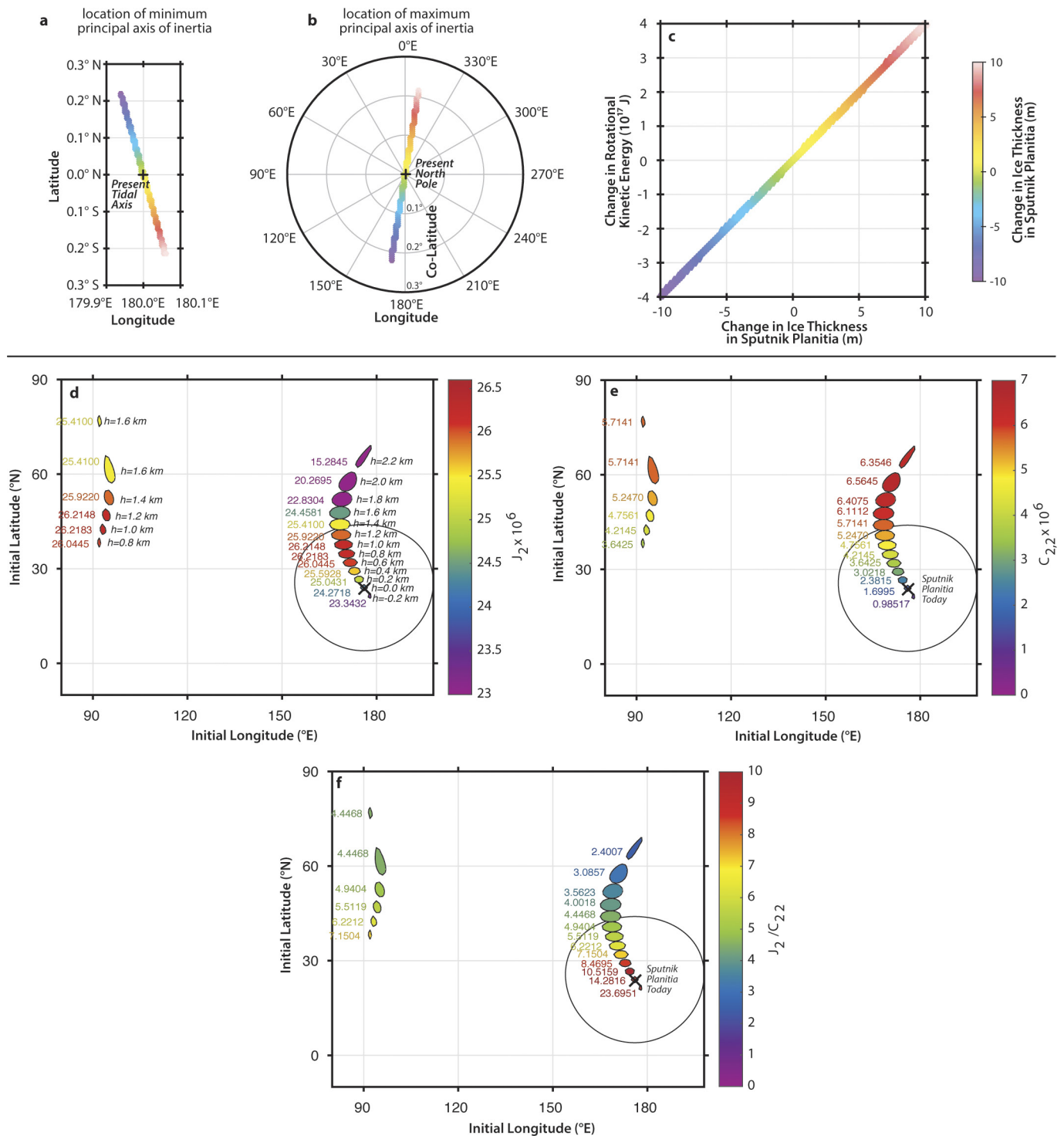


Extended Data Figure 6 | Tectonics due to the progressive loading of Sputnik Planitia. a–h, The predicted tectonic pattern on Pluto as a function of the amount of ice within Sputnik Planitia, as in Fig. 3d–i: f denotes the fraction of the total global reservoir of ice within Sputnik Planitia (where the reservoir is equivalent to a 200-m-thick global layer); $f=9\%$ corresponds to 500 m, $f=12\%$ corresponds to 800 m (as in Fig. 3f), $f=21\%$ corresponds to 1,400 m (as in Fig. 3g) and $f=52\%$ corresponds to 3,500 m. The left column shows a view above the faults west of Sputnik Planitia, while the right column shows a view above the faults east of Sputnik Planitia. Grid lines and coloured vectors denote the instantaneous principal axis reference frame. As Sputnik Planitia is loaded with volatiles,

Pluto reorients. This changes the location of these features with respect to the principal axis reference frame (resulting in the reorientation of the latitude and longitude grid in each frame). This also changes the stresses as a function of loading within Sputnik Planitia, resulting in time-evolution of the tectonic patterns. The transition from quasi-radial (loading-dominated) to quasi-azimuthal (TPW-dominated) faults increases in distance from Sputnik Planitia as a function of the loading within Sputnik Planitia. This change in stress pattern may be reflected in future study of the cross-cutting relationships of Pluto's faults. These images are snapshots from Supplementary Videos 1–3, which show additional time-steps.



Extended Data Figure 7 | Probability of randomly being near a principal axis. The probability of a randomly located feature being located near to any given principal axis (or set of axes). N is the number of axes considered; for example, the blue line indicates the chance probability of being near any single principal axis, whereas the yellow line indicates the chance probability of being near any of the principal axes. There is a 9% probability of Sputnik Planitia being this close to the tidal axis of Pluto.



Extended Data Figure 8 | Pluto's wobbles. a, b, The location of Pluto's minimum (a) and maximum (b) principal axes of inertia with respect to their present locations as a function of changes in ice thickness in Sputnik Planitia. Around 1 m of volatile ice can be transported seasonally across the entire surface of Pluto⁷, although it is unclear how volatiles migrate into and out of Sputnik Planitia. **c,** The change in rotational kinetic energy

resulting from the transport of volatile ice into and out of Sputnik Planitia. **d–f,** The spherical harmonic degree-2 gravity coefficients associated with Sputnik Planitia and the remnant figure for each of the possible reorientation scenarios shown in Fig. 2a. Measurements of degree-2 gravity (or equivalently the moments of inertia) of Pluto will constrain possible reorientation scenarios.

Reorientation of Sputnik Planitia implies a subsurface ocean on Pluto

F. Nimmo¹, D. P. Hamilton², W. B. McKinnon³, P. M. Schenk⁴, R. P. Binzel⁵, C. J. Bierson¹, R. A. Beyer⁶, J. M. Moore⁶, S. A. Stern⁷, H. A. Weaver⁸, C. B. Olkin⁷, L. A. Young⁷, K. E. Smith⁶ & the New Horizons Geology, Geophysics & Imaging Theme Team*

The deep nitrogen-covered basin on Pluto, informally named Sputnik Planitia, is located very close to the longitude of Pluto's tidal axis¹ and may be an impact feature², by analogy with other large basins in the Solar System^{3,4}. Reorientation^{5–7} of Sputnik Planitia arising from tidal and rotational torques can explain the basin's present-day location, but requires the feature to be a positive gravity anomaly⁷, despite its negative topography. Here we argue that if Sputnik Planitia did indeed form as a result of an impact and if Pluto possesses a subsurface ocean, the required positive gravity anomaly would naturally result because of shell thinning and ocean uplift, followed by later modest nitrogen deposition. Without a subsurface ocean, a positive gravity anomaly requires an implausibly thick nitrogen layer (exceeding 40 kilometres). To prolong the lifetime of such a subsurface ocean to the present day⁸ and to maintain ocean uplift, a rigid, conductive water–ice shell is required. Because nitrogen deposition is latitude-dependent⁹, nitrogen loading and reorientation may have exhibited complex feedbacks⁷.

The Sputnik Planitia basin is 3.5 km below its surroundings (Fig. 1) and is filled with a convecting layer of nitrogen ice, thought to be about 3–10 km thick^{10,11}. This structure would yield a strongly negative gravity anomaly (Extended Data Fig. 1a); to generate the present-day positive gravity anomaly either a much thicker nitrogen (N₂) layer or some other source of extra mass at depth would be required.

Stereo topography^{1,2} suggests a present-day elliptical shape of 1,300 km × 900 km. The topography resembles that of other

large degraded impact basins such as Hellas (on Mars)³ or Caloris (on Mercury)⁴ and includes a sharp rim (informally known as Cousteau Rupes) to the northeast¹. The elevated topography beyond the basin rim might represent ejecta, but a distinct ejecta blanket is not visible in images¹, perhaps because of modification by Pluto's ongoing surface geological activity. The centre of the Sputnik Planitia ellipse is at about 175° E, 18° N, or about 400 km from the tidal axis. A point randomly placed on the surface has only a 5% chance of being this close or closer to either tidal axis.

If Sputnik Planitia formed during an impact then its initial depth d_0 was probably about 7 km (see Methods), on the basis of the depths of unrelaxed basins on Iapetus and the Moon¹², with uncertainties introduced by the low velocities of Pluto impactors¹³. The horizontal scale of Sputnik Planitia suggests that a thickness of tens of kilometres of ice was removed during impact, and that impact-driven uplift of an ice–ocean interface (if the ocean is present) probably occurred¹⁴. This uplift is important because it represents a large mass excess (Extended Data Fig. 1c). On the Moon a combination of impact-driven uplift of dense mantle material and later surface addition of lavas after the crust has cooled and strengthened results in impact basins showing a positive gravity anomaly^{15–17}. We argue below that an analogous set of processes occurred at Sputnik Planitia.

If Sputnik Planitia represents a positive gravity anomaly, tidal and rotational torques will have reoriented it towards the tidal axis. The calculated reorientation is mainly equatorward (Fig. 1c) and depends on the amplitude of the positive gravity anomaly, parameterized by

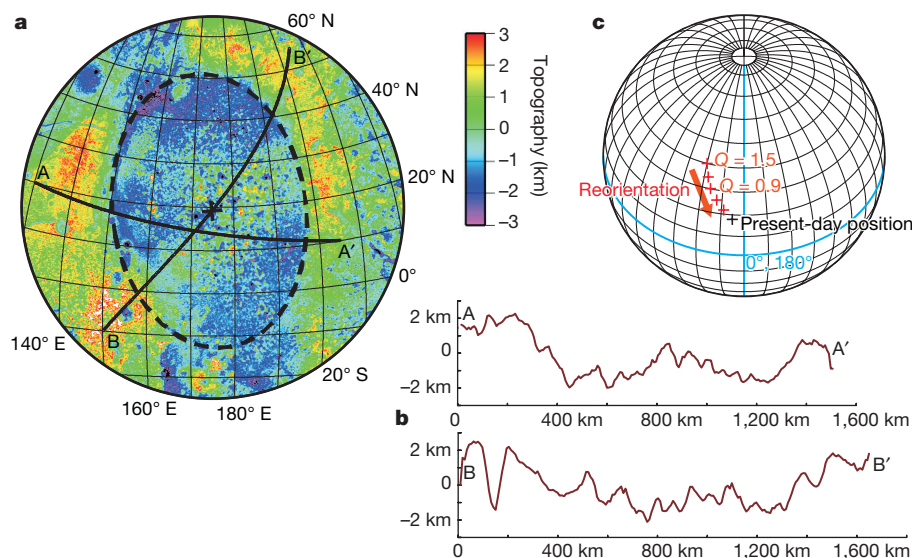


Figure 1 | Sputnik Planitia topography and reorientation. **a**, Stereo-derived topography of Sputnik Planitia (using method described in ref. 1) with an ellipse with axes 1,300 km × 900 km superimposed. The ellipse centre and projection centre (Lambert equal area) are both at 175° E, 18° N. **b**, Topographic profiles (locations shown in **a**). Point spacing was 8 km with 5-point averaging to reduce noise. **c**, Location of Sputnik Planitia before reorientation (red crosses) as a function of dimensionless gravity anomaly Q (in increments of 0.3). A Q value of 1.4 represents a nominal peak gravity anomaly Δg of +31 mGal (Methods) and yields about 20° true polar wander. Orthographic projection centred at 180° E, 45° N.

¹Department of Earth and Planetary Sciences, University of California Santa Cruz, Santa Cruz, California 95064, USA. ²Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. ³Department of Earth and Planetary Sciences and McDonnell Center for the Space Sciences, Washington University in St Louis, St Louis, Missouri 63130, USA. ⁴Lunar and Planetary Institute, Houston, Texas 77058, USA. ⁵Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁶National Aeronautics and Space Administration (NASA) Ames Research Center, Moffett Field, California 94035, USA. ⁷Southwest Research Institute, Boulder, Colorado 80302, USA. ⁸Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723, USA.

*A list of participants and their affiliations appears at the end of the paper.

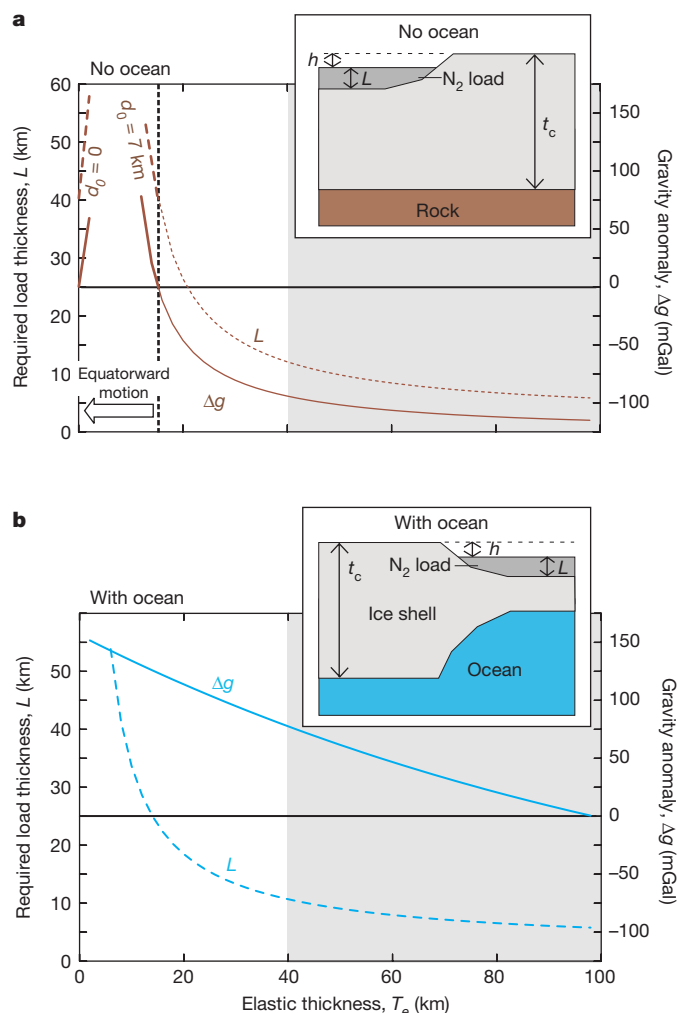


Figure 2 | Load thicknesses L and resulting gravity anomalies Δg for present-day Sputnik Planitia topography. **a**, Case with no ocean. Equatorward reorientation takes place if Δg is positive. Initial basin depth is d_0 ; to generate the present-day negative topography ($h = 3.5$ km) the deflection due to load thickness L is calculated using a thin-spherical-shell approach³⁰ (see Methods). The shaded region denotes the estimated elastic thickness range¹⁸. The characteristic wavenumber of Sputnik Planitia is taken to be $(4/3)\pi/D$, where D is the diameter (1,000 km). **b**, Case with ocean in which the pre-loading basin is isostatically compensated. Here $d_0 = 7$ km. The shell thickness t_c is taken to be $2T_e$ to calculate the gravity contribution of the water; this can be justified *a posteriori* by the requirement for a cold, conductive shell (Fig. 3). The insets show the model geometry assumed in each case.

the dimensionless parameter Q (Methods). Because of Pluto's slow spin rate, the stabilizing effect of any remnant rotational bulge is small and equatorward reorientation can occur for modest (a few tens of milligals) positive gravity anomalies. A 20° reorientation increases the probability to 23% of Sputnik Planitia's initial location being as close to a tidal axis as it is observed to be. Our calculations are conservative because they neglect the role of the ejecta blanket, silicates contained in the impactor, and decoupling of the shell from the silicates underneath, all of which will serve to increase reorientation (Methods). Conversely, if Sputnik Planitia represents a present-day negative gravity anomaly it must have formed closer to the equator (Methods).

We now calculate likely gravity anomalies at Sputnik Planitia. If no ocean was present, uplift of the silicate interior is unlikely to have happened because of its rigidity and great depth¹⁴ (assuming a differentiated body). In this case, we assume that deposition of N_2 of thickness L took place at a later epoch, by which time the crust had an

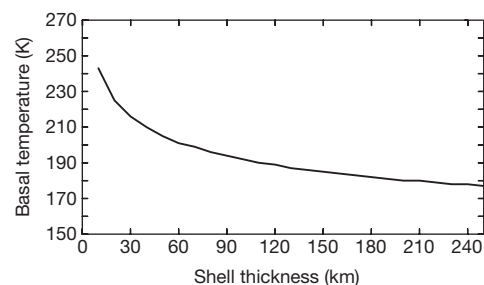


Figure 3 | Basal shell temperature required to maintain a thinned shell for 4 billion years. Timescale calculated using ref. 19 assuming a Newtonian viscosity of 10^{14} Pa s at 270 K and an activation energy of 50 kJ mol^{-1} (Methods). A conductive temperature profile was assumed with the surface at 40 K.

elastic thickness T_e . Thermal evolution models predict that T_e always exceeds 40 km, depending mainly on when Sputnik Planitia formed¹⁸. Given d_0 and the present-day topography h , the load thickness L and the resulting gravity anomaly Δg can be calculated (Fig. 2a; Methods). For basins with initial depths in the range 0–7 km, positive gravity anomalies only occur with N_2 loads over 40 km thick and T_e values less than 15 km (so that the space required by the N_2 can be accommodated). A 40-km N_2 thickness is much larger than that inferred^{10,11} and the T_e value is smaller than predicted¹⁸. The large negative gravity anomaly generated by the present-day 3.5-km negative topography is hard to overcome with N_2 loading alone.

If a subsurface ocean is present, the post-impact, pre-loading state is assumed to be isostatic, resulting in a thinned shell beneath the basin^{14,16}. The dense water beneath the basin thus provides an additional positive contribution to the overall gravity. For example, Fig. 2b shows that in the presence of an ocean, an N_2 layer 7 km thick can generate a gravity anomaly of +32 mGal for $T_e = 70$ km. These values are consistent with the available constraints.

If Sputnik Planitia is a positive gravity anomaly at the present day, Fig. 2 suggests that a subsurface ocean with a thinned shell beneath the basin provides a viable explanation. Such a configuration will be smoothed out by lateral flow of the ice¹⁹ at a rate dependent on the ice viscosity and the shell thickness t_c . Figure 3 shows that the configuration can be maintained for 4 billion years as long as the base of the ice shell is cold, that is, 180–250 K, depending on shell thickness. Such low temperatures can be achieved with an ocean containing ammonia and/or methanol²⁰ (ammonia is present in the Pluto system²¹) and imply a conductive shell, a large fraction of which will behave elastically. A conductive shell also transfers heat sufficiently slowly that a subsurface ocean can survive to the present day^{8,22}. Preferential refreezing of the thinned portion of the shell could remove shell thickness contrasts. However, the thinned portion is capped by solid N_2 , which has a much lower thermal conductivity than ice²³ and—even if convecting¹⁰—can provide sufficient insulation to prevent the thinned shell from refreezing (Methods).

Rather than uplift of liquid water underlying the ice shell, (1) uplift of mantle material, (2) dense, solid ice II, (3) silicate-rich ice or (4) reduced-porosity ice might instead be contributing to Δg . We argued above that the first possibility was unlikely. We do not favour the second alternative because the presence of ice II implies strongly compressional tectonics^{20,22}, for which there is no evidence¹. Theoretical models²⁴ predict that silicate-rich ice, if present, should be found at the surface, because of the low temperatures, while deeper ice should be silicate-free. This is opposite to the required distribution. An impact-induced porosity reduction of 10% would need to extend to a depth of 70 km to compensate the basin, but for basins the size of Sputnik Planitia the porosity effect on gravity is probably overwhelmed by uplift of the underlying material^{14,25}. Although impact-driven ocean uplift is expected for a Sputnik-Planitia-forming impact¹⁴, further work will be required to exclude these other alternatives definitively.

An alternative hypothesis²⁶ suggests that the Sputnik Planitia basin formed by early loading of N₂ ice and reorientation as Pluto's spin state evolved to synchronous. In this hypothesis N₂ was subsequently removed from Sputnik Planitia; this removal would cause >10° of polewards motion (Methods) and affect N₂ deposition. This prediction of polewards motion is opposite to that shown in Fig. 1; since reorientation²⁷ and load removal cause tectonic stresses, mapping of tectonic features⁷ should be able to test which of these hypotheses is correct.

If Pluto contains a cold (probably ammonia-bearing) liquid ocean, several further issues arise. The predicted slow re-freezing of a Plutonian ocean results in isotropic extensional stresses^{8,22}, in agreement with the tectonic features observed¹. The requirement for shell thinning to have occurred allows numerical models to probe the present-day shell thickness¹⁴. A rigid, conductive shell could be reconciled with putative cryovolcanic surface features¹ by appealing to ocean pressurization caused by progressive thickening of the ice shell²⁸. Various Kuiper Belt Objects of somewhat similar sizes and densities (bulk compositions) to Pluto are known²⁹; among these bodies, subsurface oceans are probably a common phenomenon.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 May; accepted 3 October 2016.

Published online 16 November 2016.

- Moore, J. M. *et al.* The geology of Pluto and Charon through the eyes of New Horizons. *Science* **351**, 1284–1293 (2016).
- Schenk, P. M. *et al.* A large impact origin for Sputnik Planum and surrounding terrains, Pluto? *Div. Planet. Sci. Meet.* **47**, abstr. 200.06 (2015).
- Searls, M. L. *et al.* Utopia and Hellas basins, Mars: twins separated at birth. *J. Geophys. Res.* **111**, E08005 (2006).
- Zuber, M. T. *et al.* Topography of the northern hemisphere of Mercury from MESSENGER laser altimetry. *Science* **336**, 217–220 (2012).
- Rubincam, D. P. Polar wander on Triton and Pluto due to volatile migration. *Icarus* **163**, 469–478 (2003).
- Nimmo, F. & Matsuyama, I. Reorientation of icy satellites by impact basins. *Geophys. Res. Lett.* **34**, L19203 (2007).
- Keane, J. T., Matsuyama, I., Kamata, S. & Steckloff, J. K. Reorientation and faulting of Pluto due to volatile loading within Sputnik Planitia. *Nature* <http://dx.doi.org/10.1038/nature20120> (2016).
- Robuchon, G. & Nimmo, F. Thermal evolution of Pluto and implications for surface tectonics and a subsurface ocean. *Icarus* **216**, 426–439 (2011).
- Binzel, R. P. *et al.* Climate zones on Pluto and Charon. *Icarus* <http://dx.doi.org/10.1016/j.icarus.2016.07.023> (2016).
- McKinnon, W. B. *et al.* Convection in a volatile nitrogen-ice-rich layer drives Pluto's geological and atmospheric vigour. *Nature* **534**, 82–85 (2016).
- Trowbridge, A. J. *et al.* Vigorous convection as the explanation for Pluto's polygonal terrain. *Nature* **534**, 79–81 (2016).
- White, O. L., Schenk, P. M. & Dombard, A. J. Impact basin relaxation on Rhea and Iapetus and relation to past heat flow. *Icarus* **223**, 699–709 (2013).
- Bray, V. J. & Schenk, P. M. Pristine impact crater morphology on Pluto—expectations for New Horizons. *Icarus* **246**, 156–164 (2015).
- Johnson, B. C., Bowling, T. J., Trowbridge, A. J. & Freed, A. M. Formation of the Sputnik Planum basin and the thickness of Pluto's subsurface ocean. *Geophys. Res. Lett.* **43**, 10068–10077 (2016).
- Muller, P. M. & Sjogren, W. L. Mascons – lunar mass concentrations. *Science* **161**, 680–684 (1968).
- Melosh, H. J. *et al.* The origin of lunar mascon basins. *Science* **340**, 1552–1555 (2013).
- Wieczorek, M. A. & Phillips, R. J. Lunar multiring basins and the cratering process. *Icarus* **139**, 246–259 (1999).
- Kamata, S. & Nimmo, F. Impact basin relaxation as a probe for the thermal history of Pluto. *J. Geophys. Res.* **119**, 2272–2289 (2014).
- Nimmo, F. Non-Newtonian topographic relaxation on Europa. *Icarus* **168**, 205–208 (2004).
- McKinnon, W. B., Simonelli, D. P. & Schubert, G. in *Pluto and Charon* (eds Stern, S. A. & Tholen, D. J.) 295–346 (Univ. Arizona Press, 1997).

- Grundy, W. M. *et al.* Surface compositions across Pluto and Charon. *Science* **351**, aad9189 (2016).
- Hammond, N. P., Barr, A. C. & Parmentier, E. M. Recent tectonic activity on Pluto driven by phase changes in the ice shell. *Geophys. Res. Lett.* **43**, 6775–6782 (2016).
- Scott, T. A. Solid and liquid nitrogen. *Phys. Rep.* **27**, 89–157 (1976).
- Rubin, M. E., Desch, S. J. & Neveu, M. The effect of Rayleigh-Taylor instabilities on the thickness of undifferentiated crust on Kuiper Belt Objects. *Icarus* **236**, 122–135 (2014).
- Milbury, C. *et al.* Preimpact porosity controls the gravity signature of lunar craters. *Geophys. Res. Lett.* **42**, 9711–9716 (2015).
- Hamilton, D. P. *et al.* The rapid formation of Sputnik Planitia early in Pluto's history. *Nature* <http://dx.doi.org/10.1038/nature20586> (2016).
- Matsuyama, I. & Nimmo, F. Rotational stability of tidally deformed planetary bodies. *J. Geophys. Res.* **112**, E11003 (2007).
- Manga, M. & Wang, C.-Y. Pressurized oceans and the eruption of liquid water on Europa and Enceladus. *Geophys. Res. Lett.* **34**, L07202 (2007).
- Brown, M. E. The compositions of Kuiper Belt Objects. *Annu. Rev. Earth Planet. Sci.* **40**, 467–494 (2012).
- Turcotte, D. L. *et al.* Role of membrane stresses in the support of planetary topography. *J. Geophys. Res.* **86**, 3951–3959 (1981).

Acknowledgements New Horizons was built and operated by the Johns Hopkins Applied Physics Laboratory in Laurel, Maryland, USA, for NASA. We thank the many engineers, flight controllers and others who have contributed to the success of the New Horizons mission and NASA's Deep Space Network for a decade of excellent support to New Horizons. We thank B. Johnson for discussions on impact physics and J. Conrad for cryovolcanism calculations.

Author Contributions D.P.H. originated the reorientation hypothesis. F.N. developed the subsurface ocean scenario and carried out the bulk of the calculations. C.J.B. calculated the effect of realistic basin geometries and ejecta blanket. P.M.S. and R.A.B. provided the stereo topography. All authors read or commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.N. (fnimmo@es.ucsc.edu).

Reviewer Information *Nature* thanks G. Collins and the other anonymous reviewer(s) for their contribution to the peer review of this work.

New Horizons Geology, Geophysics & Imaging Theme Team

J. M. Moore¹, W. B. McKinnon², J. R. Spencer³, R. Beyer¹, R. P. Binzel⁴, M. Buie³, B. Buratti⁵, A. Cheng⁶, D. Cruikshank¹, C. Dalle Ore¹, A. Earle⁴, R. Gladstone⁷, W. Grundy⁸, A. D. Howard⁹, T. Lauer¹⁰, I. Linscott¹¹, F. Nimmo¹², J. Parker³, S. Porter³, H. Reitsema¹³, D. Reuter¹⁴, J. H. Roberts⁵, S. Robbins³, P. M. Schenk¹⁵, M. Showalter¹⁶, K. Singer³, D. Strobel¹⁷, M. Summers¹⁸, L. Tyler¹¹, O. L. White¹, O. M. Umurhan¹, M. Banks¹⁹, O. Barnouin⁶, V. Bray²⁰, B. Carcich³, A. Chaikin²¹, C. Chavez², C. Conrad³, D. P. Hamilton²², C. Howett³, J. Hofgartner²³, J. Kammer³, C. Lisse⁶, A. Marcotte⁶, A. Parker³, K. Retherford³, M. Saina⁶, K. Runyon⁵, E. Schindhelm³, J. Stansberry²⁴, A. Steffl³, T. Stryck²⁵, H. Throop³, C. Tsang³, A. Verbiscer⁹, H. Winters⁶, A. Zangari³, S. A. Stern³, H. A. Weaver⁶, C. B. Olkin³, L. A. Young³ & K. E. Smith¹

¹National Aeronautics and Space Administration (NASA) Ames Research Center, Moffett Field, California 94035, USA. ²Department of Earth and Planetary Sciences and McDonnell Center for the Space Sciences, Washington University in St Louis, Saint Louis, Missouri 63130, USA. ³Southwest Research Institute, Boulder, Colorado 80302, USA. ⁴Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁵NASA Jet Propulsion Laboratory, Pasadena, California 91019, USA. ⁶Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723, USA. ⁷Southwest Research Institute, San Antonio, Texas 78238, USA. ⁸Lowell Observatory, Flagstaff, Arizona 86001, USA. ⁹University of Virginia, Charlottesville, Virginia 22904, USA. ¹⁰National Optical Astronomy Observatory, Tucson, Arizona 85719, USA. ¹¹Stanford University, Stanford, California 94305, USA. ¹²Department of Earth and Planetary Sciences, University of California Santa Cruz, Santa Cruz, California 95064, USA. ¹³B612 Foundation, Mill Valley, California 94941, USA. ¹⁴NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ¹⁵Lunar and Planetary Institute, Houston, Texas 77058, USA. ¹⁶The SETI Institute, Mountain View, California 94043, USA. ¹⁷The Johns Hopkins University, Baltimore, Maryland 21218, USA. ¹⁸George Mason University, Fairfax, Virginia 22030, USA. ¹⁹Planetary Science Institute, Tucson, Arizona 85719, USA. ²⁰University of Arizona, Tucson, Arizona 85721, USA. ²¹Arlington, Vermont 05250, USA. ²²University of Maryland, College Park, Maryland 20742, USA. ²³Cornell University, Ithaca, New York 14853, USA. ²⁴Space Telescope Science Institute, Baltimore, Maryland 21218, USA. ²⁵Roane State Community College, Oak Ridge, Tennessee 37830, USA.

METHODS

Reorientation. To calculate the reorientation caused by Sputnik Planitia loading we follow the methods of ref. 27 with one exception. For a tidally distorted, slowly rotating synchronous satellite, the ratio of the non-normalized hydrostatic degree-two gravity coefficients is $J_2/C_{22} = 10/3$. However, since Pluto is the primary, it experiences less tidal distortion and the coefficient ratio is correspondingly higher, about 14.3 (ref. 7). As a result, we generalize equation (39) of ref. 27 as follows:

$$Q \sin 2\theta_L \cos(\phi_R - \phi_L) = \sin 2\theta_R \left(1 + \frac{f \cos^2 \theta_T}{\sin^2 \theta_R} \right) \quad (1)$$

Here Q is the dimensionless load size, θ and ϕ are colatitude and longitude, respectively, and the subscripts L, T and R refer to the final location of the load and the initial locations of the tidal axis and the rotational axis in the final reference frame. Here f is defined as $f = 3m/(M+m)$, where m and M are the masses of the tide-raising body (Charon) and Pluto, respectively, such that for a synchronous satellite orbiting a massive planet $f = 3$ (yielding equation (39) of ref. 27) while for a purely rotationally distorted body $f = 0$. With this modification the reorientation caused by an imposed load Q may be calculated. For simplicity, we assume that reorientation occurs as a single event, though in reality it may have consisted of progressive motion.

For $Q = 1$ and $f = 0.327$ (appropriate to Pluto) we find that $\theta_T = 102.5^\circ$, $\phi_T = 193.1^\circ$, $\theta_R = 13.6^\circ$ (this is the amount of true polar wander, TPW) and $\phi_R = 169.6^\circ$. A TPW of 20° requires $Q = 1.4$. The initial position of the load in the final reference frame may then be calculated using spherical triangles or by diagonalizing the moment of inertia tensor (ref. 27); for $Q = 1$ the load is initially located at 31.6° latitude and 163.2° longitude in the final reference frame.

A basin of constant depth h and angular radius ψ yields the following dimensionless load Q (ref. 6):

$$Q = \frac{3\pi G h \rho \cos \psi \sin^2 \psi}{R \Omega^2 \Delta k_2} = \frac{3}{2} \frac{p \Delta g \cos \psi \sin^2 \psi}{R \Omega^2 \Delta k_2} \quad (2)$$

where G is the gravitational constant, ρ is the density of the material, R and Ω are the radius and rotation angular frequency of Pluto and Δk_2 is the difference between the fluid Love number and the actual Love number (this quantity describes the size of the remnant bulge, which opposes reorientation). The numerator depends on the size of the load and the denominator represents the remnant bulge size. The size of the remnant bulge depends on Δk_2 and the rotation rate at which the bulge was 'frozen in'. Existing shape observations show no evidence of a remnant bulge³¹ and the establishment of Pluto's present-day spin rate probably took a few million years³¹, whereas cooling of the interior and freezing in of a remnant bulge probably took tens to hundreds of millions of years^{8,22}. We therefore take the relevant rotation rate to be that of the present day. The second equality introduces the peak gravity anomaly Δg associated with the basin. For a parabolic basin (as we assume for Sputnik Planitia), the peak gravity is the same as for the constant-depth case, but the corresponding value of Q is reduced by a factor of $p \approx 0.5$ because the mean basin depth is smaller. We take $R = 1,188$ km, $p = 0.5$, $\Omega = 1.14 \times 10^{-5}$ rad s⁻¹, and $\psi = 24^\circ$ ($D = 1,000$ km). For a Pluto with a 50-km-thick elastic lithosphere $\Delta k_2 = 0.16$ (see below) in which case equation (2) yields $\Delta g = 22Q$ mGal. A larger Δk_2 (larger remnant bulge) would require a larger gravity anomaly to get the same amount of reorientation.

Our calculated degree of reorientation is probably conservatively small, for three reasons. First, if present, an ejecta blanket will reduce the size of the original negative gravity anomaly associated with the basin (yielding $p \approx 0.3$). Second, the basin-forming impactor probably contained some silicates, so any impactor material incorporated into the ice shell will provide a positive contribution to gravity. Third, a decoupled ice shell is likely to reorient more than a solid body. However, for our argument the degree of reorientation is less important than the sign: only a basin exhibiting a positive gravity anomaly will experience equatorward reorientation.

Polewards motion. For a load near the tidal axis and for a body (like Pluto) that is primarily rotationally distorted, we can approximate equation (1) as $Q \sin 2\theta_L \approx \sin 2\theta_R$ with $\theta_L = 72^\circ$ for present-day Sputnik Planitia. The present-day gravity anomaly in the absence of a subsurface ocean is about -115 mGal (Extended Data Fig. 1a). Using the present-day rotation period and setting $\Delta k_2 = 1$ to represent the largest likely remnant bulge (the real value is probably considerably smaller; see below) and with $\Delta g = -115$ mGal, equation (2) shows that the corresponding value of Q is -0.8 . This in turn implies a poleward reorientation θ_R of about 14° , and an original (pre-reorientation) latitude of 4° . A smaller remnant bulge would result in more reorientation. If Sputnik Planitia is a negative gravity anomaly at the present day, or if mass was removed after its equilibrium

position was established, Sputnik Planitia should have experienced large poleward reorientation, because the stabilizing effect of the rotational remnant bulge is small.

Loading calculations. Consider first a basin that is initially isostatically compensated by an uplifted root (the with-ocean case), so that the initial gravity anomaly is about zero. The initial uplift r is given by $r = d_0 \rho_c / (\rho_m - \rho_c)$, where ρ_m and ρ_c are the density of water and ice, respectively, and d_0 is the depth of the basin after rebound. Assuming that an initially unstressed elastic layer develops after the rebound is complete, subsequent loading results in deflection. Taking the load thickness to be L , the deflection w (positive downwards) and the final basin negative topography h , we have

$$h = d_0 + w - L \quad (3)$$

For a load described by a single spherical harmonic degree n , the required load thickness L for a given h can then be obtained via

$$L = \frac{(h - d_0)(C'_n + 1)}{\left(\frac{\rho_L - \rho_c}{\rho_c} C'_n - 1 \right)} \quad (4)$$

Here ρ_L is the load density, $C'_n = \frac{\rho_c}{\rho_m - \rho_c} C_n$ where C_n is the degree of compensation³⁰, which depends on the elastic thickness, and we have modified the definition from ref. 30 to avoid singularities arising when $\rho_m = \rho_c$. In the rigid limit there is no deflection, so $C'_n = 0$, and equation (4) yields the correct answer: $L = d_0 - h$. In the isostatic limit $C'_n = \frac{\rho_c}{\rho_m - \rho_c}$ and again the correct answer is recovered, yielding a much larger load thickness: $L = (d_0 - h) \rho_m / (\rho_m - \rho_L)$.

The post-loading peak gravity anomaly is given by

$$\Delta g = 2\pi G(-[h + L]\rho_c + L\rho_L + [r - w][\rho_w - \rho_c]e^{-kt_c}) \quad (5)$$

The final term in equation (5) represents the positive gravity contribution of the uplifted dense water. Here the factor $\exp(-kt_c)$ is due to upwards attenuation of the gravity anomaly owing to the finite shell thickness t_c . We take $t_c = 2T_c$.

Next we consider a basin overlying a flat ice-silicate interface (the no-ocean case). The depth after any initial (pre-loading) flexure is taken to be d_0 . The required load thickness can again be obtained from equation (4), where in this case C'_n is calculated by setting $\rho_m = \rho_c$ (because there is no contribution from a higher-density layer at depth). Again, the correct answer is recovered in the rigid and isostatic limiting cases. In this case the peak gravity anomaly is then simply

$$\Delta g = 2\pi G(-[h + L]\rho_c + L\rho_L) \quad (6)$$

We calculate C_n using equation (27) of ref. 30. We convert from wavenumber k to spherical harmonic degree n by using $n \approx kR$. The Young's modulus of ice is 9 GPa, densities of water ice, water and N₂ ice are taken to be 0.92 g cm⁻³, 1.0 g cm⁻³ and 1.0 g cm⁻³ (ref. 23), respectively. Incorporation of ammonia into the ice could in theory reduce its effective rigidity, but during slow freezing ammonia will be excluded from the crystallizing ice³².

In reality, Sputnik Planitia loading consists of contributions from multiple wavenumbers. To determine the dominant wavenumber, we calculated the flexural deflection of a parabolic basin using the approach of ref. 33 and determined that the maximum deflection is well approximated by an effective wavenumber $k = 4\pi/3D$, where D is the basin diameter.

Lateral flow of the shell. The timescale for lateral flow of the shell is calculated using the approach of ref. 22, which gives the relaxation timescale τ :

$$\tau = \frac{\eta_b}{g \Delta \rho \delta^3 k^2}$$

where η_b is the basal viscosity, k is the wavenumber as before, δ is the effective layer thickness in which flow occurs and $\Delta \rho$ is the ice-water density contrast. The basal viscosity depends on the reference viscosity and the activation energy Q_a , and for a shell in which conductivity varies as $1/T$, δ is given by

$$\delta = \frac{R_g T_b t_c}{Q_a \ln(T_b/T_s)}$$

where R_g is the gas constant and T_b and T_s are the basal and surface temperatures. **Size of remnant bulge.** The size of the remnant bulge^{27,34} is assumed to depend on the quantity $k_{2f} - k_2$, where k_{2f} is the Love number after all stresses have relaxed and k_2 is the present-day Love number. A body which is fluid at the present day has

no remnant bulge ($k_{2f} - k_2 = 0$) while a body which is infinitely rigid now ($k_2 = 0$) has the largest possible remnant bulge, the size of which depends on the density structure and initial rotation rate. We use the method of ref. 35 to calculate the Love numbers and assume that the body is spherically symmetric. We assume that Pluto's silicate interior has remained rigid and unrelaxed at all timescales and has an outer radius of 842 km, a rigidity of 100 GPa and a density of 3.5 g cm^{-3} . The overlying H_2O layer has a mean density of 0.95 g cm^{-3} and an outer radius of 1,188 km. In the presence of an elastic ice shell 50 km thick with a shear modulus of 3 GPa, $k_2 = 0.28$, while in the absence of such a shell $k_{2f} = 0.44$. The fact that $(k_{2f} - k_2) \approx k_2$ implies that the remnant bulge and present-day bulge are of comparable magnitude. Our assumption of a rigid silicate core is based on thermal evolution calculations⁸; if the core were instead strengthless at all timescales, the Love numbers increase to $k_2 = 0.52$ and $k_{2f} = 0.75$, respectively.

Initial depth of Sputnik Planitia basin. Pluto's radius is close to the geometric mean of the radii of Iapetus ($R = 734 \text{ km}$) and the Moon ($R = 1,738 \text{ km}$). Basins a thousand kilometres across that appear to be unrelaxed exist on Iapetus and the Moon¹², with Iapetus basins approaching 10 km in depth and lunar basins about a factor of two shallower. A similar-scale unrelaxed basin on Pluto might therefore be expected to be about 7 km deep. The corresponding isostatic ocean uplift would be 80 km. Expected impact velocities on Pluto are lower even than on Iapetus, but the implications of these lower velocities for the initial depth-to-diameter ratio of the resulting basin are unclear¹³.

The extent to which crust (shell) thinning and mantle (ocean) uplift occur in response to an impact depend on the diameter of the basin relative to the depth to the mantle, or ocean^{14,16}. On the Moon, with a mean crustal thickness of about 35 km, mantle uplift occurs for basins with diameters in excess of 220 km (refs 25 and 36). Assuming that this same ratio applies to Pluto, a 1,000-km-diameter basin would be expected to generate ocean uplift for shells thinner than about 160 km. This expectation is confirmed by numerical models¹⁴, which show that uplift occurs for ice-shell thicknesses less than about 180 km. A chondritic Pluto might have a present-day shell thickness similar to this value^{8,24}, while in the past the shell will have been thinner and uplift will correspondingly be more likely to have occurred.

Insulating effect of N_2 . Consider a reference shell of thickness t_c and effective thermal conductivity k_c . It may be compared with a thinned shell of total thickness t'_c containing a layer of lower-conductivity ice k' of thickness L . For the heat fluxes

across the two shells to be equal, the required thickness of the insulating ice L can be shown to be

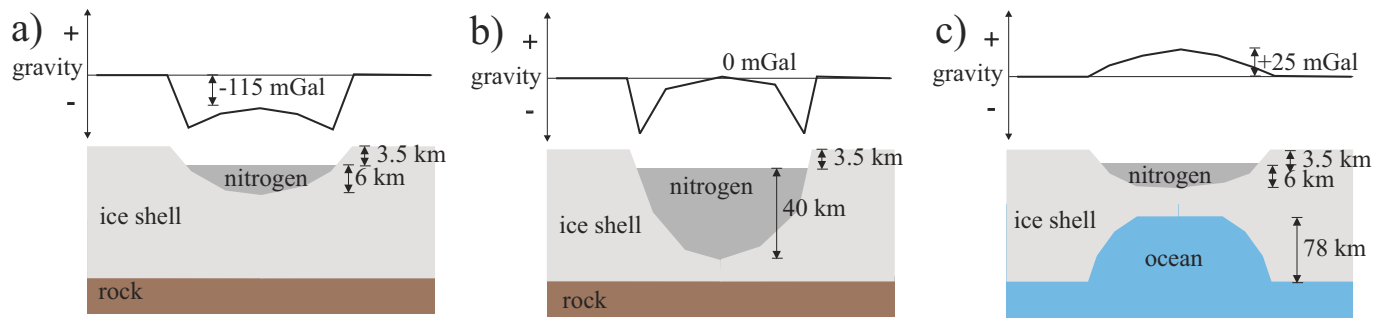
$$L = \frac{k'}{k_c - k'}(t_c - t'_c)$$

We note that this analysis neglects any melting at the base of the N_2 layer. Water ice exhibits a temperature-dependent thermal conductivity given by $651/T$ (ref. 37). The effective thermal conductivity k_c over the temperature range 40–240 K is then $5.8 \text{ W m}^{-1} \text{ K}^{-1}$. In contrast, nitrogen ice at 50 K has a thermal conductivity of $0.2 \text{ W m}^{-1} \text{ K}^{-1}$ (ref. 23). The effective thermal conductivity of the nitrogen will be increased if it is convecting. Based on the results of ref. 10, the Nusselt number of the convecting nitrogen is about 3, so that the effective nitrogen thermal conductivity $k' \approx 0.6 \text{ W m}^{-1} \text{ K}^{-1}$.

For an initial basin depth of 7 km, the shell thinning after loading ($t_c - t'_c$) at Sputnik Planitia will be about 70 km, depending on the exact densities assumed and the amount of deformation. Thus, a nitrogen layer 8 km thick is sufficient to offset the increased heat flux due to the thinned shell. As a result, shell thickness variations can be maintained over geological timescales as long as an insulating N_2 ice layer persists. As shown in Fig. 2, a layer this thick will yield a positive gravity anomaly of about +30 mGal, sufficient to cause reorientation.

Code availability. Codes for the reorientation, loading and lateral flow calculations are available upon request from F.N.

31. Nimmo, F. *et al.* Mean radius and shape of Pluto and Charon from New Horizons images. *Icarus* <http://dx.doi.org/10.1016/j.icarus.2016.06.027> (2016).
32. Kargel, J. S. Ammonia water volcanism on icy satellites—phase relations at 1-atmosphere. *Icarus* **100**, 556–574 (1992).
33. Comer, R. P., Solomon, S. C. & Head, J. W. Mars—thickness of the lithosphere from the tectonic response to volcanic loads. *Rev. Geophys.* **23**, 61–92 (1985).
34. Willemann, R. J. Reorientation of planets with elastic lithospheres. *Icarus* **60**, 701–709 (1984).
35. Moore, W. B. & Schubert, G. The tidal response of Europa. *Icarus* **147**, 317–319 (2000).
36. Soderblom, J. M. *et al.* The fractured Moon: production and saturation of porosity in the lunar highlands from impact cratering. *Geophys. Res. Lett.* **42**, 6939–6944 (2015).
37. Petrenko, V. F. & Whitworth, R. W. *Physics of Ice* (Clarendon Press, 1999).



Extended Data Figure 1 | Schematic of the way in which the gravity anomaly is affected by an uplifted ocean and the thickness of the nitrogen layer. a–c, Either a nitrogen layer more than 40 km thick (b) or an uplifted ocean (c) could result in the present-day positive gravity anomaly at Sputnik Planitia; if neither is present, then a negative gravity anomaly results (a). The peak gravity anomaly is calculated using the flat-plate formula $2\pi G\Delta\rho h$ for each layer, where h represents the thickness,

$\Delta\rho$ is the lateral density contrast and the densities of H_2O ice, water and N_2 ice are 0.92 g cm^{-3} , 1.0 g cm^{-3} and 1.0 g cm^{-3} (ref. 23), respectively. In c, the gravitational contribution of the ocean is reduced as a result of upwards attenuation assuming a shell thickness of 150 km (see Methods). The structure in c is similar to the inferred structure of lunar mascon basins, which also show positive gravity anomalies (refs 15, 16).

The rapid formation of Sputnik Planitia early in Pluto's history

Douglas P. Hamilton¹, S. A. Stern², J. M. Moore³, L. A. Young² & the New Horizons Geology, Geophysics & Imaging Theme Team*

Pluto's Sputnik Planitia is a bright, roughly circular feature that resembles a polar ice cap. It is approximately 1,000 kilometres across and is centred on a latitude of 25 degrees north and a longitude of 175 degrees, almost directly opposite the side of Pluto that always faces Charon as a result of tidal locking¹. One explanation for its location includes the formation of a basin in a giant impact, with subsequent upwelling of a dense interior ocean². Once the basin was established, ice would naturally have accumulated there³. Then, provided that the basin was a positive gravity anomaly (with or without the ocean), true polar wander could have moved the feature towards the Pluto–Charon tidal axis, on the far side of Pluto from Charon^{2,4}. Here we report modelling that shows that ice quickly accumulates on Pluto near latitudes of 30 degrees north and south, even in the absence of a basin, because, averaged over its orbital period, those are Pluto's coldest regions. Within a million years of Charon's formation, ice deposits on Pluto concentrate into a single cap centred near a latitude of 30 degrees, owing to the runaway albedo effect. This accumulation of ice causes a positive gravity signature that locks, as Pluto's rotation slows, to a longitude directly opposite Charon. Once locked, Charon raises a permanent tidal bulge on Pluto, which greatly enhances the gravity signature of the ice cap. Meanwhile, the weight of the ice in Sputnik Planitia causes the crust under it to slump, creating its own basin (as has happened on Earth in Greenland⁵). Even if the feature is now a modest negative gravity anomaly, it remains locked in place because of the permanent tidal bulge raised by Charon. Any movement of the feature away from 30 degrees latitude is countered by the preferential recondensation of ices near the coldest extremities of the cap. Therefore, our modelling suggests that Sputnik Planitia formed shortly after Charon did and has been stable, albeit gradually losing volume, over the age of the Solar System.

The western lobe of the icy, heart-shaped feature on Pluto, provisionally called Sputnik Planitia (Extended Data Fig. 1), is composed of a mixture of nitrogen, methane and carbon monoxide ices^{1,6,7}—substances that are volatile at the temperatures expected on Pluto (about 40 K). Its location at 30° N—a latitude that is temperate on Earth—currently receives substantially less solar energy per Pluto year than all other latitudes. Furthermore, the axial tilt and orbital eccentricity of Pluto evolve on million-year timescales^{8,9}, leading to important changes in the annual energy flux to the dwarf planet and to climate variations that are analogous to ice-age cycles on Earth. We investigate these variations^{10–19} by considering the instantaneous flux of sunlight that is absorbed by a unit surface element on Pluto:

$$F = \frac{L}{4\pi r^2} (1 - A) \cos(\gamma) \quad (1)$$

where L is the solar luminosity, r is the Pluto–Sun distance, A is Pluto's albedo and γ is the angle between the Sun–Pluto line and the normal to the surface element.

We average equation (1) over one orbit of Pluto about the Sun (see Methods and Extended Data Fig. 2), assuming a perfect absorber ($A = 0$) to obtain the solar energy flux to different latitudes on Pluto. Pluto's current axial tilt (or obliquity) is $\epsilon = 120^\circ$, but it varies from

104° to 127° over the course of a 2.8-million-year cycle⁹. Accordingly, we repeat the calculation for several different obliquities (Fig. 1). The curve for the present-day tilt has a minimum flux at 26.5° latitude, with more solar energy incident at the equator and substantially more at the poles. In the current epoch, Pluto's poles receive about 10% more solar energy than does its equator, but this ratio varies from about 45% more at $\epsilon = 104^\circ$ to near parity at $\epsilon = 127^\circ$. Volatiles are driven away from the poles during the long hot summers, preferentially migrating to regions with cooler climates¹². Once deposited, these volatiles are more difficult to dislodge, resulting in thinner layers returning to the poles in winter. Over many repeated seasonal cycles, these processes inexorably move volatiles away from the poles and towards cooler regions. Current conditions near the equator, and especially near 30°, strongly favour the deposition and retention of volatiles.

Over the 2.8-million-year cycle of Pluto's obliquity, the equator is alternately more and less hospitable to the build up of ices than is 30° latitude. The flow of ices across the equator, as evidenced in Extended Data Fig. 1, might be due to the recent warming trend at the equator, which could be mobilizing the ices in the southernmost regions of Sputnik Planitia. Most importantly, conditions near 30° are remarkably constant over Pluto's full obliquity cycle, providing a stable climate in which to cultivate icy deposits over millions and even billions of years. Weaker effects of Pluto's orbital eccentricity and radiation from Charon are considered in Methods.

If incident sunlight was the only factor in determining the location of frozen volatiles on Pluto, then planet-encircling bands of ices centred near $\pm 30^\circ$ would be expected. However, the surface albedo of the planet determines the fraction of incident energy that is actually absorbed by Pluto. Dark areas in Extended Data Fig. 1 absorb about 85% of the incident sunlight ($A = 0.15$ in equation (1)), whereas the bright ices of Sputnik Planitia reflect a comparable fraction of incident sunlight back into space ($A \approx 0.85$)⁶. Owing to diminished absorption of sunlight, temperatures over bright icy regions should be much lower than over darker regions, thereby enhancing the deposition rate of volatiles and inhibiting sublimation¹⁷. The strong effect of albedo can drive longitudinal variations in ice cover. The dark equatorial regions of Pluto remain dark by discouraging deposition of bright frost layers, whereas the bright areas attract additional frost. The equally marked albedo variations on Saturn's moon Iapetus are driven by similar processes²⁰. Acting in concert with variations in incident sunlight, albedo differences will cause volatiles to be preferentially deposited on the brightest terrain in the $\pm 30^\circ$ latitude bands. Perhaps the most likely trigger for the formation of Sputnik Planitia was bright terrain formed from refreezing of melt produced by a mid-sized impactor or an early period of cryovolcanism. Shadowing by mountains or in craters are also possibilities, although these are probably weaker effects. In any case, once ices begin to be deposited in one area, they raise the albedo and reflect more sunlight back to space. Thus, the absorption of energy decreases, sublimation is inhibited and deposition of more ices is enhanced, leading to a runaway process (Fig. 2; see Methods for more details). Runaway ice-cap formation on Pluto is the reverse of the accelerating melting that is now ongoing in Earth's north polar ice pack.

¹University of Maryland, College Park, Maryland, USA. ²SWRI, Boulder, Colorado, USA. ³NASA Ames, Mountain View, California, USA.

*A list of participants and their affiliations appears at the end of the paper.

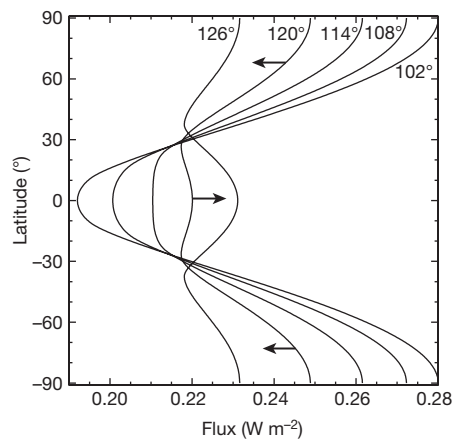


Figure 1 | Effects of Pluto's obliquity on insolation. We plot the orbit-averaged incident solar energy flux, or insolation, as a function of latitude on Pluto for five different values of Pluto's axial tilt ϵ . All curves were computed assuming the present-day value of Pluto's eccentricity, $e = 0.25$. The change in Pluto's tilt is the largest single factor that affects the strength of incident sunlight; the 1.4-million-year swing from $\epsilon = 104^\circ$ to $\epsilon = 127^\circ$ causes a change of roughly 20% in the flux absorbed at Pluto's equator and poles. Latitudes near $\pm 30^\circ$ have a far more muted response to changes in Pluto's obliquity. The black arrows highlight the present-day obliquity of Pluto and show how the obliquity will change in the future; the minimum insolation at the equator occurred about 0.85 million years ago and the peak is due in about 0.55 million years.

Owing to the strength of even a thin lithosphere, the initial ice deposits would have had positive topography contributing to a global longitudinal asymmetry. Such asymmetry would be subject to tidal torques from Charon that would act to rotate the ice cap towards the Pluto–Charon line. This probably occurred very early, within a few million years after Pluto and Charon were formed by the impact of two large Kuiper belt objects²¹. Prior to this giant impact, the surfaces of each of the objects were probably cold and covered with frozen volatiles and thin atmospheres. The collision produced either a mostly intact Charon or a circum-Pluto disk out of which Charon subsequently accreted; it also substantially heated portions of each body²¹. Tides acting between Pluto and Charon then despun each object and separated them on a timescale of the order of a few million years^{9,22–24}.

Far shorter timescales are sufficient for surface temperatures of Pluto to come back into equilibrium with the flux of solar energy and, accordingly, for most of Pluto's volatiles to freeze out onto its surface. Deposition would occur preferentially at latitudes near the equator if Pluto's obliquity then was below 114° , or nearer latitudes of $\pm 30^\circ$ for larger obliquities (Fig. 1), and at longitudes with favourably high albedos. Over many seasonal cycles of sublimation and deposition, the runaway albedo effect (discussed above and in Methods) will cause a single ice cap to form in at most a few hundred thousand years. As Charon approached its current distance from Pluto, the obliquity of the dwarf planet stabilized to something like its current value, and Charon locked into synchronous orbit above a particular spot on Pluto's surface. The ice cap was plausibly the largest contributor to global topography (see Methods), dwarfing in horizontal scale the mountains and craters scattered across Pluto's surface. In this circumstance, Charon would cease moving relative to Pluto's surface when it reached one of two equilibrium longitudes: that of the ice cap itself or 180° further on (see Fig. 3 and Methods). For this type of tidal alignment, all that is needed is for the ice cap to control the longitudinal asymmetry of Pluto when averaged over broad 90° swaths of longitude. On these large scales, the gravitational signature from the mass of the ice cap exceeds the integrated effect of Pluto's other topography. The vertical and horizontal extent of the ice cap is also relatively unimportant, and it is sufficient for the ice to initially simply fill in topographical lows. The runaway albedo effect, acting over time, probably concentrated the ice into the deep convecting structure that we now observe (see Methods). The current

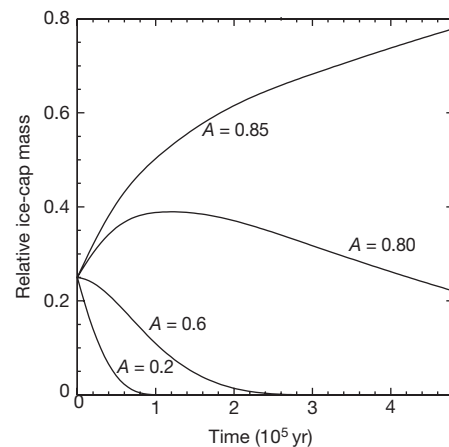


Figure 2 | The runaway albedo effect. This simulation tracks the time evolution of four ice caps with initially equal mass, each with an initial radius of 800 km and an average depth of 500 m, but with different albedos A . In reality, one feature probably dominated immediately after ices first started condensing out of the atmosphere, leading to much faster formation than is depicted here. Nevertheless, in each seasonal cycle of 248 years, we assume that a fraction $f = 0.5$ of absorbed solar energy converts ice to vapour, which is then redeposited on the ice caps in proportion to their exposed surface areas. As ice caps gain or lose mass, their areal coverage and average depth evolve as well, such that the ratio of diameter to depth remains constant. The darker two ice caps ($A = 0.2$ and $A = 0.6$) efficiently absorb solar energy and sublimate away rapidly, while the brighter two ($A = 0.8$ and $A = 0.85$) initially grow. Eventually, the brighter two ice caps compete more directly for the available vapour, and the one with the higher albedo captures all of the available ice, growing to a depth of 700 m and a radius of 1,270 km. Finally, the depth-to-diameter ratio of the surviving ice cap will grow as the lower-albedo margins preferentially sublimate in a process analogous to that considered here.

position of Sputnik Planitia, centred nearly opposite the sub-Charon longitude of 0° , provides excellent support for this idea of moon-driven tidal alignment. Indeed, if randomly placed, there is only an approximately 10% chance that the ice cap would fall within 10° of one of the two critical longitudes, 0° and 180° .

Once Charon comes to rest above Pluto, the gravity of the satellite sets up a permanent tidal bulge on the dwarf planet that reinforces the position of the ice cap at a longitude that intersects the Pluto–Charon line. This permanent bulge holds the ice cap in place even against substantial mass loss of volatiles from the ice cap to the atmosphere, space or elsewhere on the planet. The ice cap would rotate away from 180° longitude only if enough ice was moved to reveal a large negative gravity anomaly, or basin, and the basin would have to be sufficiently large and deep to overcome the permanent tidal bulge. Therefore, once locked in, the ice cap will remain near its current equilibrium longitude of 180° against all but the most extreme changes. Moderately sized ice caps would also move slightly towards the equator^{2,4,25}, but such motions are resisted by Pluto's rotational bulge. Furthermore, any equatorward displacement of the ice cap, or indeed even an initially equatorial ice cap, would be affected by many thousands of annual sublimation and deposition cycles that would move the ice cap slowly poleward towards the latitudes with the least orbit-averaged flux of sunlight (see Methods). The north–south orientation of Sputnik Planitia may provide some evidence for such motions. Eventually conditions stabilized, leaving Pluto with a single dominant ice cap centred on a latitude of 25° determined by a minimum of solar illumination and a longitude of 175° set by long-ago tidal forces from Charon. Therefore, Pluto should be in one of four possible end states: with an ice cap centred near 30° or -30° latitude and near 0° or 180° longitude. Positions within 10° of these end states cover just 3% of Pluto's surface, making Sputnik Planitia's location near one of them particularly noteworthy.

The ices within Sputnik Planitia contain solid-state convection cells that suggest depths of several kilometres^{1,26,27}. Convection works to

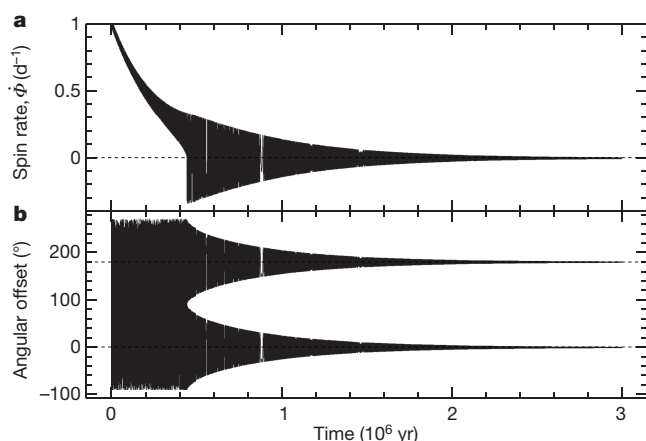


Figure 3 | Tidal locking of Pluto. **a**, The spin rate of Pluto relative to Charon's orbital motion ($\dot{\Phi}$) is shown for two separate initial conditions differing in phase by 180° . The two curves are plotted on top of one another (see Methods). Torques from Charon on the tidal bulges that are raised on Pluto by the moon act to slow Pluto's relative spin rate towards zero, where it ultimately locks to Charon's orbital rate ($\dot{\Phi} = 0$, indicated by the dashed line) to form the doubly synchronous system that we observe today. **b**, The longitudinal offset between Charon and the ice cap on Pluto's surface is shown for the two initial conditions. After a time of rapid rotation, each simulated Pluto is captured into libration about an equilibrium point, one with the longitudes of Pluto's ice cap and Charon aligned and the other with them separated by 180° ; these critical longitudes are marked by dashed lines.

keep the surface bright by continually exposing fresh ice and reprocessing dirty ice, protecting the ice cap against additional absorption of sunlight. Despite its thickness, the surface of Sputnik Planitia is actually a topographic low relative to the global average, residing in a large basin²⁶. This depression has been interpreted as having an impact origin^{2,26,28}—a scenario that is not wholly inconsistent with our arguments; such a structure might, for example, be created by the loss shortly after its formation of a hypothetical moon similar to the four that currently reside outside Charon's orbit^{1,29}. However, there is a simpler alternative: mass loading of Pluto's surface with a layer of ice many kilometres thick early in Pluto's history when its rigid lithosphere was thin would depress the surface substantially; the ice cap would make its own basin. This interpretation has the advantage of providing an explanation for why the basin is coincident with the ice cap and why both are located at the coldest latitude on Pluto and at a longitude that is directly opposite Charon. Because nitrogen ice and the water-ice bedrock have comparable densities, the nitrogen ice cap would have efficiently depressed the pre-existing landscape. In this case, the fact that the volume of the basin substantially exceeds that of Sputnik Planitia itself²⁶ implies that some of the ices that once filled the basin are now elsewhere, possibly lost to space over billions of years^{30,31}, hidden in the unlit southern hemisphere or buried underneath equatorial or other deposits. In addition, bulk movement of the ices on million-year precessional timescales may have eroded the regions around Sputnik Planitia, further enlarging the basin.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 May; accepted 28 October 2016.

1. Stern, S. A. *et al.* The Pluto system: initial results from its exploration by New Horizons. *Science* **350**, aad1815 (2015).
2. Nimmo, F. *et al.* Reorientation of Sputnik Planitia implies a subsurface ocean on Pluto. *Nature* <http://dx.doi.org/10.1038/nature20148> (2016).
3. Bertrand, T. & Forget, F. Observed glacier and volatile distribution on Pluto from atmosphere-topography processes. *Nature* <http://dx.doi.org/10.1038/nature19337> (2016).

4. Keane, J. T., Matsuyama, I., Kamata, S. & Steckloff, J. K. Reorientation and faulting of Pluto due to volatile loading within Sputnik Planitia. *Nature* <http://dx.doi.org/10.1038/nature20120> (2016).
5. Spada, G. *et al.* Greenland uplift and regional sea level changes from ICESat observations and GIA modelling. *Geophys. J. Int.* **189**, 1457–1474 (2012).
6. Grundy, W. M. *et al.* Surface compositions across Pluto and Charon. *Science* **351**, aad9189 (2016).
7. Protopapa, S. *et al.* First look at global Pluto's surface composition through pixel-by-pixel radiative scattering model of New Horizons Ralph/LEISA data. *Icarus* (in the press); preprint at <https://arxiv.org/abs/1604.08468> (2016).
8. Dobrovolskis, A. R. & Harris, A. W. The obliquity of Pluto. *Icarus* **55**, 231–235 (1983).
9. Dobrovolskis, A. R., Peale, S. J. & Harris, A. W. in *Pluto and Charon* (eds Stern, S. A. & Tholen, D. J.) 159–190 (Univ. Arizona Press, 1997).
10. van Hemelrijck, E. The insolation at Pluto. *Icarus* **52**, 560–564 (1982).
11. van Hemelrijck, E. Insolation changes on Pluto caused by orbital element variations. *Earth Moon Planets* **33**, 163–177 (1985).
12. Binzel, R. P. 1991 Urey prize lecture: physical evolution in the solar system—present observations as a key to the past. *Icarus* **100**, 274–287 (1992).
13. Spencer, J. R. *et al.* in *Pluto and Charon* (eds Stern, S. A. & Tholen, D. J.) 435–473 (Univ. Arizona Press, 1997).
14. Lissauer, J. J., Barnes, J. W. & Chambers, J. E. Obliquity variations of a moonless Earth. *Icarus* **217**, 77–87 (2012).
15. Earle, A. M. & Binzel, R. P. Pluto's insolation history: latitudinal variations and effects on atmospheric pressure. *Icarus* **250**, 405–412 (2015).
16. Hamilton, D. P. The icy cold heart of Pluto. In *47th Division for Planetary Sciences Meeting abstr.* **200.07** (American Astronomical Society, 2015).
17. Earle, A. M. *et al.* Long-term surface temperature modeling of Pluto. *Icarus* <http://dx.doi.org/10.1016/j.icarus.2016.09.036> (2016).
18. Binzel, R. P. *et al.* Climate zones on Pluto and Charon. *Icarus* <http://dx.doi.org/10.1016/j.icarus.2016.07.023> (2016).
19. Nadeau, A. & McGhee, R. A. A simple method for calculating a planet's mean annual insolation by latitude. Preprint at <https://arxiv.org/abs/1510.04542> (2015).
20. Spencer, J. R. & Denk, T. Formation of Iapetus' extreme albedo dichotomy by exogenically triggered thermal ice migration. *Science* **327**, 432–435 (2010).
21. Canup, R. M. A giant impact origin of Pluto-Charon. *Science* **307**, 546–550 (2005).
22. Peale, S. J. in *Planetary Satellites* (ed. Burns, J. A.) 87–111 (Univ. Arizona Press, 1977).
23. Murray, C. D. & Dermott, S. F. *Solar System Dynamics* **165** (Cambridge University Press, 2000).
24. Cheng, W. H., Lee, M. H. & Peale, S. J. Complete tidal evolution of Pluto-Charon. *Icarus* **233**, 242–258 (2014).
25. Matsuyama, I., Mitrovica, J. X., Manga, M., Perron, J. T. & Richards, M. A. Rotational stability of dynamic planets with elastic lithospheres. *J. Geophys. Res. Planets* **111**, E02003 (2006).
26. Moore, J. M. *et al.* The geology of Pluto and Charon through the eyes of New Horizons. *Science* **351**, 1284–1293 (2016).
27. McKinnon, W. B. *et al.* Convection in a volatile nitrogen-ice-rich layer drives Pluto's geological vigour. *Nature* **534**, 82–85 (2016).
28. Johnson, B. C., Bowling, T. J., Trowbridge, A. J. & Freed, A. M. Formation of the Sputnik Planum basin and the thickness of Pluto's subsurface ocean. *Geophys. Res. Lett.* **43**, 10068–10077 (2016).
29. Showalter, M. R. & Hamilton, D. P. Resonant interactions and chaotic rotation of Pluto's small moons. *Nature* **522**, 45–49 (2015).
30. Gladstone, G. R. *et al.* The atmosphere of Pluto as observed by New Horizons. *Science* **351**, aad8866 (2016).
31. Bagenal, F. *et al.* Pluto's interaction with its space environment: solar wind, energetic particles, and dust. *Science* **351**, aad9045 (2016).

Acknowledgements We thank NASA for its support of the New Horizons mission and the New Horizons mission team for making the July 2015 flyby possible. We thank V. Bray, B. Carcich, J. Hofgartner and F. Nimmo for helpful comments. This research was supported by a grant from NASA Origins (to D.P.H.).

Author Contributions D.P.H. wrote the manuscript and the computer codes used to produce all of the figures. S.A.S., L.A.Y. and J.M.M. commented on draft manuscripts and have leadership roles with New Horizons that helped make the mission possible.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.P.H. (dphamil@umd.edu).

New Horizons Geology, Geophysics & Imaging Theme Team

R. P. Binzel¹, M. W. Buie², B. J. Buratti³, A. F. Cheng⁴, K. Ennico⁵, W. M. Grundy⁶, I. R. Linscott⁷, W. B. McKinnon⁸, C. B. Olkin², H. J. Reitsema⁹, D. C. Reuter¹⁰, P. Schenk¹¹, M. R. Showalter¹², J. R. Spencer², G. L. Tyler⁷ & H. A. Weaver⁴

¹MIT, Cambridge, Massachusetts, USA. ²SWRI, Boulder, Colorado, USA. ³JPL, Pasadena, California, USA. ⁴Applied Physics Lab, Laurel, Maryland, USA. ⁵NASA Ames, Mountain View, California, USA. ⁶Lowell Observatory, Flagstaff, Arizona, USA. ⁷Stanford University, Stanford, California, USA. ⁸Washington University, St Louis, Missouri, USA. ⁹Ball Aerospace, Denver, Colorado, USA. ¹⁰NASA Goddard, Greenbelt, Maryland, USA. ¹¹Lunar and Planetary Institute, Houston, Texas, USA. ¹²SETI Institute, Mountain View, California, USA.

METHODS

Solar energy deposition on Pluto. We begin by averaging the energy flux from sunlight to a given latitude l_y on Pluto over the course of one Pluto rotation. The integration is over time with P_{day} equal to one Pluto day or approximately 6.39 Earth days. The daily average flux to a given latitude is given by the integral of equation (1) with $A = 0$:

$$\bar{F}_{\text{day}} = \frac{L}{4\pi r^2} \frac{1}{P_{\text{day}}} \int_0^{P_{\text{day}}} \cos(\gamma) dt = \frac{L}{4\pi r^2} \frac{1}{\pi} \int_0^{\beta_{\text{max}}} \cos(\gamma) d\beta \quad (2)$$

where L is the solar luminosity and r is the instantaneous Pluto–Sun distance, which is assumed to remain approximately constant over one Pluto day. The angle γ is the angular distance between the Sun and the point of interest on Pluto's surface (Extended Data Fig. 2); γ changes continuously in time owing to Pluto's rotation. We have used the fact that β —the hour angle measured from the north pole (Extended Data Fig. 2)—increases uniformly as Pluto rotates so that $\beta/(2\pi) = t/P_{\text{day}}$. Finally, β_{max} is the largest β such that the Sun is still visible ($\gamma = 90^\circ$). Equation (2) shows that the daily average flux is highest when both β_{max} and $\cos(\gamma)$ are maximized (long sunlit periods with the Sun nearly overhead); this situation is approximately realized at two locations along the orbit of Uranus where that planet's spin axis points nearly at the Sun.

The integral is evaluated by expanding the angle γ with the law of cosines for the spherical triangle shown in Extended Data Fig. 2. We obtain the following result^{32,33}:

$$\bar{F}_{\text{day}} = \frac{L}{4\pi r^2} \frac{1}{\pi} [\beta_{\text{max}} \sin(l_s) \sin(l_y) + \sin(\beta_{\text{max}}) \cos(l_s) \cos(l_y)] \quad (3)$$

where l_s is the subsolar latitude, l_y is the latitude of interest and $\cos(\beta_{\text{max}}) = -\tan(l_s) \tan(l_y)$, with $\beta_{\text{max}} \in [0, \pi]$. For circular orbits about the Sun, note the symmetries $\bar{F}_{\text{day}}(l_s, l_y) = \bar{F}_{\text{day}}(-l_s, -l_y)$ and $\bar{F}_{\text{day}}(l_s, l_y) = \bar{F}_{\text{day}}(l_y, l_s)$. Equation (3) can be used to show that, on the summer solstice, Pluto's summer pole receives nearly 5.5 times the energy than its equator does. In fact, Pluto receives more daily sunlight at its north pole than at its equator whenever $l_s > 17.66^\circ$. For higher latitudes, the more direct sunlight to the equator is more than offset by the greater amount of time that the polar regions spend in sunlight. Of course, the equator is sunlit for exactly half a Pluto rotation at all times of the year. For a given position of the Sun north of the equator, the greatest daily flux of energy is always to a more northerly latitude. For example, the daily flux to latitude $l_y = 17.66^\circ$ when $l_s = 17.66^\circ$ is less than that to all latitudes l_y between 17.66° and 41.8° .

To determine the annually averaged energy flux, we integrate equation (3) over a full Pluto year:

$$\bar{F}_{\text{year}} = \frac{L}{4\pi} \frac{1}{P_{\text{year}}} \int_0^{P_{\text{year}}} \frac{1}{r^2} [\beta_{\text{max}} \sin(l_s) \sin(l_y) + \sin(\beta_{\text{max}}) \cos(l_s) \cos(l_y)] dt$$

We use Kepler's second law

$$\frac{d\nu}{dt} = \frac{h}{r^2} \quad (4)$$

to replace integration over time with an integration over the angle ν . Here $h = \sqrt{GM_{\odot}a(1-e^2)}$ is Pluto's constant orbital angular momentum per unit mass, a , e and ν are Pluto's orbital semimajor axis, eccentricity and true anomaly, respectively, G is the gravitational constant and M_{\odot} is the solar mass. The true anomaly is the angle between Pluto and its orbital pericentre as measured from the Sun. Note that the r^2 introduced by this change of variables conveniently cancels the r^{-2} from the definition of flux. We find

$$\bar{F}_{\text{year}} = \frac{L}{8\pi^3} \frac{1}{a^2 \sqrt{1-e^2}} \int_0^{2\pi} [\beta_{\text{max}} \sin(l_s) \sin(l_y) + \sin(\beta_{\text{max}}) \cos(l_s) \cos(l_y)] d\nu \quad (5)$$

Together, the true anomaly ν , its time derivative (equation (4)) and the argument of latitude u defined by

$$\sin(u) = \sin(\epsilon) \sin(l_s) \quad (6)$$

give full information about the apparent motion of the Sun as seen from Pluto. The argument of latitude is the angle between the present-day position of the Sun and its position when it crosses Pluto's equator heading north; ϵ is Pluto's obliquity. The true anomaly and argument of latitude are related by $u = w + \nu$, where w is the argument of pericentre; $w = 0$ occurs at times when Pluto is closest to the Sun at its spring equinox. Because the integrand is a complicated function of the true anomaly, the integral in equation (5) must be evaluated numerically.

Although intractable, the integral displays some interesting symmetries. First, taking $\epsilon \rightarrow \pi - \epsilon$ leaves equation (6) unaltered and so $\bar{F}_{\text{year}}(\epsilon, l_y) = \bar{F}_{\text{year}}(\pi - \epsilon, l_y)$. Consequently, Pluto, with a tilt of $\epsilon = 120^\circ$, gets the same annual insolation pattern on its surface as a planet tilted by $\epsilon = 60^\circ$. More remarkably, if we rewrite the equation as an integral over u using $du = d\nu$, the limits of integration cover a full cycle of the Sun around Pluto and can be returned to the range $[0, 2\pi]$. Therefore, the integral does not depend in any way on the location of Pluto's pericentre⁹. The only effect of Pluto's highly eccentric orbit on \bar{F}_{year} comes from the overall scaling factor outside the integral³³. This means that the annual average energy flux to any latitude on Pluto is most easily determined by treating its orbit as circular for the integration in equation (5), and then adjusting the scaling factor outside the integral (see Extended Data Fig. 3).

A final symmetry comes from flipping the subsolar point and the latitude of interest to the opposite hemisphere with the transformation $l_s \rightarrow -l_s$ and $l_y \rightarrow -l_y$. This flipping occurs when the Sun is half an orbit further along ($u \rightarrow u + \pi$), so $du \rightarrow du$ and we rewrite the integration limits as before. Therefore, $\bar{F}_{\text{year}}(\epsilon, l_y) = \bar{F}_{\text{year}}(\epsilon, -l_y)$ and the annual energy flux to a planet is always symmetric about its equator. This symmetry, which is clearly visible in Fig. 1, is seen in some works^{10,11,14,17,19}, but not others^{13,15}.

The solution of equation (5) shows that cold polar regions as exist on Earth or Mars require obliquities of less than 45° or more than 135° , whereas deep equatorial minima occur for all planets tilted between 66° and 114° . Pluto's obliquity varies around the 114° boundary, spending about 1.3 million years with an equatorial minimum followed by 1.5 million years with minima at low to mid latitudes (Fig. 1). **Numerical techniques.** We wrote several codes in C to calculate the energy flux from the Sun for different orbital geometries and axial tilts over various time intervals. We evaluate equation (3) directly to quickly determine average daily fluxes at different times during Pluto's year; to obtain annual averages, we evaluate equation (5). For averages over one Pluto year, millions of years in the past or future, we update Pluto's orbital parameters and tilt appropriately⁹ and again evaluate equation (5).

As seen from Pluto, Charon is an extremely large and extremely close satellite, appearing fully seven times larger than the full Moon in Earth's sky, which itself is larger than any other planetary satellite seen from its primary. Moreover, owing to its tidally evolved state, Charon hovers over one point on Pluto's equator, continuously illuminating the same hemisphere. Accordingly, we investigate the reflected sunlight and thermal infrared emission from Charon on ices on Pluto's surface. Because Charon holds its position over a single spot on Pluto's equator, the angle between this spot and the position of interest γ_C is constant. We estimate the average flux of solar energy that comes to Pluto via Charon as follows:

$$\bar{F}_{C,\text{year}} = \bar{F}_{\text{year}} \frac{1}{2} \frac{R_C^2}{a_C^2} \cos(\gamma_C) \quad (7)$$

where R_C is Charon's radius and a_C is its distance from Pluto. Here we have assumed that Charon is full as seen from Pluto, have treated Charon as a slow rotator and have made simplifying assumptions about how its surface reflects sunlight. We have made use of \bar{F}_{year} from equation (5), which introduces errors of approximately 20%. Charon's phases vary over a Pluto year, always oscillating about, but rarely departing appreciably from, half full; in the spirit of estimation, we ignore these subtleties, but recognize that our full-moon approximation is an overestimate. Additionally, we underestimate the effects of Charon's thermal emission because equation (7) determines the energy incident on Pluto rather than the energy absorbed. More than half of the sunlight incident on Charon is reprocessed into thermal infrared, which delivers energy to ices far more efficiently than does visible light. Finally, more indirect energy from Charon is delivered to the night side of Pluto than to the dayside, which probably has consequences that are not captured by our approach of considering only annual averages. We also estimate the effect of sunlight reductions to Pluto due to eclipses by Charon that occur every 124 years³⁴. When averaged over a Pluto year, this effect is smaller, but not substantially smaller, than reflected sunlight and thermal emission from Charon. With all of these caveats, our estimate for contributions of Charon shine to Pluto's sub-Charon longitude is depicted in Extended Data Fig. 3; despite Charon's size and proximity, these effects are very small—roughly an order of magnitude weaker than the already weak effects of orbital eccentricity.

This analysis shows that variations in Pluto's obliquity and albedo are the strongest drivers of sublimation and deposition, with weaker but noticeable effects from orbital eccentricity and even radiation from the large moon Charon. Although we find a slight preference for ices to be deposited on the anti-Charon hemisphere, the effect is too weak to account for the location of Sputnik Planitia.

Runaway albedo effect. To illustrate the runaway albedo effect, we developed a code to track the individual sizes of several spatially separated ice deposits through

Pluto's seasonal cycles. We assume that a fraction $1 - A$ of incident sunlight is absorbed by the ices, with A the albedo from equation (1). We further assume that a fraction f of absorbed sunlight converts ice to vapour during sunlight hours and that all of this vapour redeposits onto the ice caps at night. The rest of the incident energy goes to heating of the surface and subsurface layers, and redeposition is assumed to be in proportion to the exposed surface areas of the ice caps. Over a single summer day ice preferentially sublimates into the atmosphere, whereas in winter the converse is true. We consider only the longer annual cycle, assuming that the vapour content of the atmosphere remains constant and using it only as a conduit for communication from one ice deposit to another. We set the number of ice caps, and their sizes, depths and albedos as initial conditions, and track changes to the deposits over many annual cycles. Figure 2 shows a typical run with four large ice caps of initially equal size, but different albedos. The four-ice-cap situation in Fig. 2 is unlikely to have ever occurred on Pluto, but serves as an effective demonstration of the power of the runaway albedo effect.

Albedos are macroscopic averages, encoding the effects of large-scale surface topography as well as compositional differences. The ice caps may be placed at arbitrary locations on Pluto's surface, but here we consider caps near a configuration in which each spot receives the same annual solar insolation. The total mass in ice is set to approximate that believed to be in the ice cap today, and is preserved over the course of the simulation. We find that mass is rapidly removed from the darker two ice caps, owing to their more efficient absorption of solar radiation and consequent copious production of vapour. Simultaneously, the brighter two ice caps both grow in mass and spatial extent, with their larger surface areas attracting more of the available vapour each annual cycle. Ultimately, and on relatively short timescales, only one ice cap remains.

Although this simple model does not capture all of the relevant physics—for example, it does not include the expected brightening of the ice cap once its size exceeds that necessary for solid-state convection²⁷—it effectively demonstrates that very small albedo differences are magnified to the point where only a single ice cap survives. Similar considerations explain why an excess of solar radiation at polar and equatorial latitudes drives ices towards latitudes that receive the least solar illumination (Fig. 1).

Tidal despinning of Pluto. We developed a numerical code to track how the angular orientation of Pluto relative to Charon changes as a result of tidal influences from the massive satellite. The relative spin rate of Pluto starts positive and then slows to zero, or tidally locks, over a timescale of about a million years^{9,22–24}, where it remains today. Gravitational torques from Charon on the tidal bulge raised on Pluto by this satellite cause a monotonic decrease in the spin rate of the dwarf planet. In addition, the dominant torques on the asymmetric shape of Pluto vary periodically at twice the relative spin rate. We show below that Pluto's asymmetric shape is probably dominated by its ice cap. These two effects are captured by the following equation for the rotational acceleration of Pluto:

$$\frac{d^2\Phi}{dt^2} = \epsilon_1 \text{sign}\left(\frac{d\Phi}{dt}\right) + \epsilon_2 \sin(2\Phi) \quad (8)$$

where Φ is the rotational phase of Pluto relative to Charon, t is time, $d\Phi/dt$ is the relative spin rate, ϵ_1 is the tidal dissipation rate and ϵ_2 is set by the intrinsic shape of Pluto. The first term on the right-hand side of equation (8) captures tidal despinning of Pluto that occurs as a result of tides raised on that body by Charon. Here we use the so-called constant- Q formulation of tides²³, which, although in common usage, has a very obvious flaw. The tidal term takes a discontinuous step when $d\Phi/dt = 0$, which is unphysical. When $d\Phi/dt = 0$, which corresponds to the spin rate of Pluto precisely matching the orbital rate of Charon, the tidal bulge raised on Pluto by Charon is directly under the satellite and there can be no net

torque. An infinitesimally faster or slower spin rate should lead to an infinitesimal torque that drives the system back towards $d\Phi/dt = 0$, not the macroscopic torque implied by equation (8). This formulation of tides also leads to numerical issues, because no step size in time is sufficiently small to resolve a step function. Accordingly, we smooth the second term in equation (8) by making it proportional to $d\Phi/dt$ when $|d\Phi/dt|$ is less than a specific threshold. This choice is more physical, maintains continuity, is numerically more stable and has the added advantage of making equation (8) linear; in fact, for small Φ it is the equation of a classic damped harmonic oscillator. In the code, this change takes the form of a single conditional statement. Taking nominal parameters in equation (4.165) from ref. 23, we calculate $\epsilon_1 = 1.4 \times 10^{-18} \text{ s}^{-2}$ and choose our threshold spin rate to be $d\Phi/dt = 1.0 \text{ rad d}^{-1}$.

The last term in equation (8) characterizes torques from Charon on the longitudinally asymmetric shape of Pluto. Because Charon's distance is large compared to Pluto's radius, the global-scale topography (specifically the C_{22} quadrupolar component of topography) of Pluto determines its response to Charon. A single ice cap of mass m_{ice} , placed at the equator of an otherwise spherical planet, yields $C_{22} = m_{\text{ice}}/(4M_{\text{P}})$, where M_{P} is Pluto's mass; the contribution from an ice cap at 30° latitude is only 25% smaller. No dipole terms are produced by the addition of an ice cap as long as the origin of the system remains at Pluto's centre of mass. Therefore, a single ice cap of mass m_{ice} acts identically to two ice caps each of mass $m_{\text{ice}}/2$ located 180° apart in longitude—a situation that is easier to visualize. For the parameters of the ice cap considered in Fig. 2, we have $m_{\text{ice}}/M_{\text{P}} = 3 \times 10^{-4}$, $C_{22} = 8 \times 10^{-5}$ and

$$\epsilon_2 = \frac{15}{4} \frac{GM_{\text{P}}}{a_{\text{C}}^3} \frac{m_{\text{ice}}}{M_{\text{P}}} \frac{M_{\text{C}}}{M_{\text{P}}} = 1.5 \times 10^{-14} \text{ s}^{-2}$$

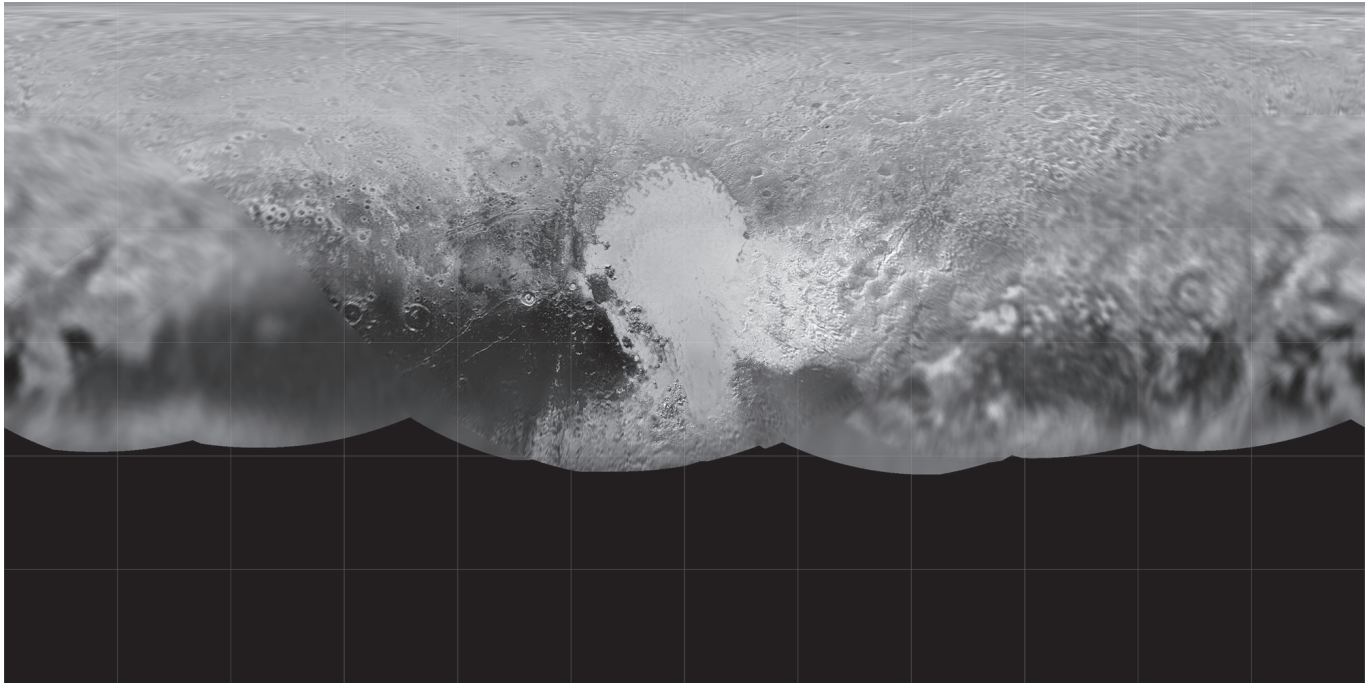
where M_{C} is Charon's mass. Pluto's natural topography, if similar to Earth's, would yield a considerably smaller $C_{22} = 2.8 \times 10^{-6}$, providing support for our contention that Pluto's ice cap probably dominated the global topography. Substantial relaxation of Pluto towards spherical symmetry can occur without altering this conclusion.

The transformation $\Phi \rightarrow \Phi + \pi$ leaves equation (8) unaltered, which means that any trajectory that Φ follows has a second solution that is everywhere displaced from the first by 180° . When we integrate equation (8) in time, we find two equilibrium states that occur with equal probability: the ice cap can end up at Charon's longitude ($\Phi = 0$) or displaced by 180° to where the actual system is found. Figure 3 follows the evolution of Pluto's spin rate relative to Charon's orbit and the angular offset of the long axis of Pluto from Charon as a function of time. Because $\epsilon_2 \gg \epsilon_1$, the harmonic oscillator is underdamped and the system oscillates around the equilibrium point, within an envelope that decays exponentially. Our numerical simulations support our analytic discussion: we find rapid decay of Pluto's spin to a final state with the ice cap locked either to Charon's longitude or to roughly the observed configuration, each with 50% probability.

Code availability. The numerical codes used to produce Fig. 1 and Extended Data Fig. 3 are not publicly available. The codes developed to illustrate the runaway albedo effect (Fig. 2) and despinning of Pluto (Fig. 3) are available from D.P.H. on request.

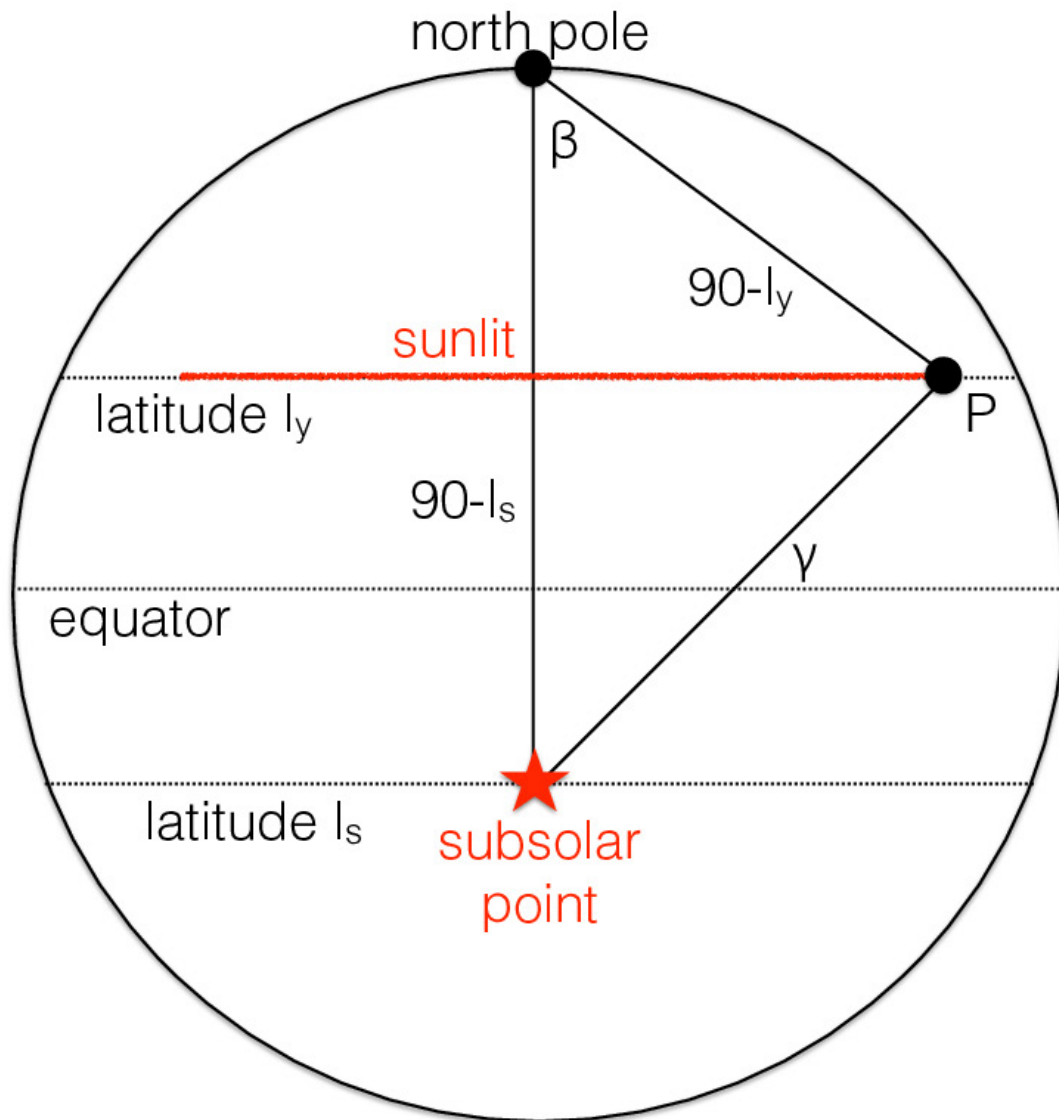
Data availability. Data files used in the construction of the figures are available from D.P.H. on request.

32. Cross, C. A. The heat balance of the Martian polar caps. *Icarus* **15**, 110–114 (1971).
33. Ward, W. R. Climatic variations on Mars: 1. Astronomical theory of insolation. *J. Geophys. Res.* **79**, 3375–3386 (1974).
34. Buie, M. W., Tholen, D. J. & Horne, K. Albedo maps of Pluto and Charon: initial mutual event results. *Icarus* **97**, 211–227 (1992).



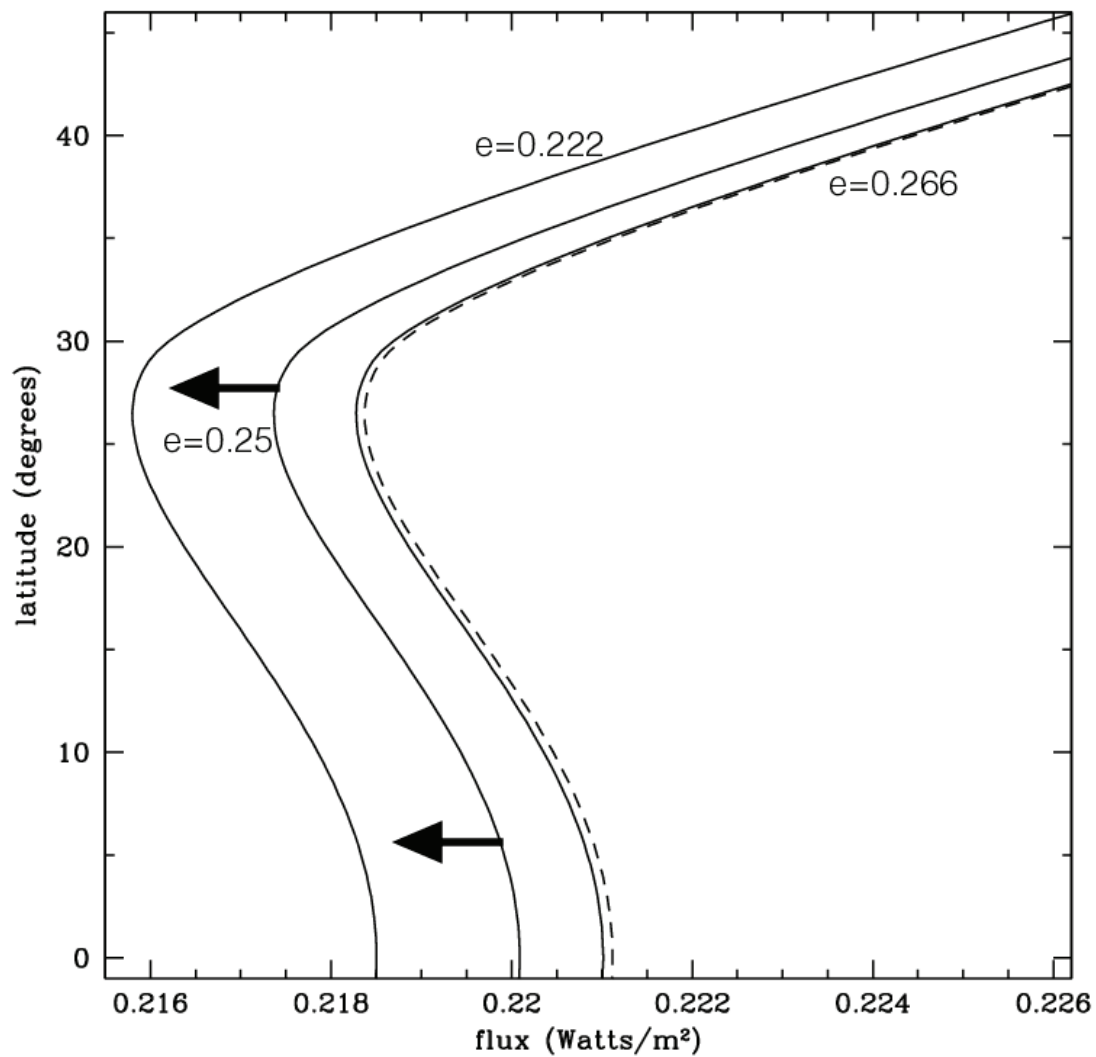
Extended Data Figure 1 | Surface map of Pluto. Sputnik Planitia, the informally named western lobe of the white heart-shaped feature, is approximately 1,000 km across and is centred on a latitude of 25° N and longitude of 175° , with the zero of longitude defined to run directly underneath Pluto's moon Charon. Owing to Pluto's slow rotation and the approach vector of New Horizons, the highest-resolution images

were of Sputnik Planitia, and most other regions were imaged at much lower resolution from farther away. Consequently, this map is a mosaic of multiple images of differing resolution. Faint grid lines of latitude and longitude are spaced by 30° . Regions south of about 30° S were not sunlit and, hence, were not imaged, because they are currently experiencing polar night.



Extended Data Figure 2 | Geometry for energy deposition on Pluto. The large circle represents Pluto, with our view centred on the intersection of the equator and the noon meridian. At the location of the red star, the Sun is directly overhead, and it tracks along the latitude l_s over the course of a full rotation of Pluto. The thick horizontal red line segment shows the regions along latitude l_y that are currently illuminated by sunlight. The angle γ is the angular distance between the Sun and the point of interest on

Pluto's surface (black dot labelled 'P' at latitude l_y), as measured from the centre of Pluto. The Sun is on the horizon when $\gamma = 90^\circ$, which we define to occur at the meridional angle of $\beta = \beta_{\max}$. The spherical triangle formed by γ and the two meridians connecting the north pole to the above-mentioned points defines γ in terms of the other variables and simplifies derivations of average energy fluxes.



Extended Data Figure 3 | Effects of Pluto's eccentricity on insolation. Here we zoom in on the region surrounding Sputnik Planitia and show (solid curves) the solar energy flux with Pluto at its minimum ($e = 0.222$), current ($e = 0.25$) and maximum ($e = 0.266$) eccentricity, assuming a present-day obliquity of 120° . Black arrows show the future changes that are expected as a result of these eccentricity variations. As we show in Methods, Pluto's eccentricity affects insolation equally at all latitudes.

Incident radiation from Charon, a 0.1% effect, moves the solid $e = 0.266$ curve slightly to the right, most noticeably at the equator, as indicated by the dashed curve and discussed in Methods. Eccentricity effects are at the 1% level, 20 times weaker than the obliquity effects highlighted in Fig. 1. The effect of Pluto's eccentricity varies with a 3.95-million-year period⁸. The maximum insolation to the entire planet occurred about 0.8 million years ago and the minimum will next occur in about 1.2 million years.

Ghost imaging with atoms

R. I. Khakimov¹, B. M. Henson¹, D. K. Shin¹, S. S. Hodgman¹, R. G. Dall¹, K. G. H. Baldwin¹ & A. G. Truscott¹

Ghost imaging is a counter-intuitive phenomenon—first realized in quantum optics^{1,2}—that enables the image of a two-dimensional object (mask) to be reconstructed using the spatio-temporal properties of a beam of particles with which it never interacts. Typically, two beams of correlated photons are used: one passes through the mask to a single-pixel (bucket) detector while the spatial profile of the other is measured by a high-resolution (multi-pixel) detector. The second beam never interacts with the mask. Neither detector can reconstruct the mask independently, but temporal cross-correlation between the two beams can be used to recover a ‘ghost’ image. Here we report the realization of ghost imaging using massive particles instead of photons. In our experiment, the two beams are formed by correlated pairs of ultracold, metastable helium atoms³, which originate from *s*-wave scattering of two colliding Bose–Einstein condensates^{4,5}. We use higher-order Kapitza–Dirac scattering^{6–8} to generate a large number of correlated atom pairs, enabling the creation of a clear ghost image with submillimetre resolution. Future extensions of our technique could lead to the realization of ghost interference⁹, and enable tests of Einstein–Podolsky–Rosen entanglement⁹ and Bell’s inequalities¹⁰ with atoms.

Since the first experimental realization of ghost imaging with light more than twenty years ago^{11,12}, following the original proposals some years previously^{13,14}, there has been much interest in the technique^{1,2}. Ghost imaging has found applications in areas including environmental sensing¹⁵ and cryptography¹⁶. Extension of these techniques has enabled ghost imaging in the X-ray domain^{17,18}, three-dimensional ghost imaging¹⁹ and even temporal ghost imaging, which could potentially lead to improved telecommunications²⁰. Ghost imaging has produced sub-shot-noise images of weakly absorbing objects²¹ and, for low-light-level imaging when the number of registered photons per image pixel is less than one, has been shown to outperform conventional imaging in terms of contrast²².

There was considerable debate as to whether ghost imaging is a semi-classical or a quantum-optical phenomenon. It has been shown that ghost imaging can be realized either using a source of thermal^{23,24} or pseudothermal²⁵ photons, or using correlated photon pairs (bi-photons) such as are created in spontaneous parametric down conversion (SPDC)²⁴. Thermal or pseudothermal ghost imaging is not a quantum effect and can be explained semi-classically, whereas SPDC ghost imaging requires quantum theory to describe some aspects of its quantitative performance and can demonstrate violation of Bell’s inequalities¹⁰. However, the key enabler of ghost imaging (whether using thermal or quantum sources) is the existence of particle correlations in the source used for the ghost imaging.

Experimental ghost imaging has so far been achieved exclusively with photons. Here we report on the experimental realization of ghost imaging using massive particles. As well as demonstrating complementarity between light and matter for this phenomenon, realizing ghost imaging with atoms is a potential precursor to experiments that test fundamental concepts in quantum mechanics with massive particles, such as ghost interference, Einstein–Podolsky–Rosen entanglement and Bell’s inequalities^{9,10,12}. Further, there could potentially be future applications such as real-time, *in situ* control

of atom lithography while imaging the deposition remotely via the spatial (multi-pixel) detector.

There are two main challenges to ghost imaging with massive particles: first, the need for a source of such particles with the required correlation properties; and second, the need for a sufficiently high flux of particles to provide adequate measurement statistics. Atom cooling techniques that enable the creation and manipulation of Bose–Einstein condensates (BECs) allow such a source to be created. A schematic of ghost imaging using pairs of correlated, colliding atoms is illustrated in Fig. 1.

As with optics, two types of atom sources can be envisaged. Like thermal light sources, a thermal ensemble of ultracold atoms possesses second- ($1 < g^{(2)} < 2$) and higher-order correlations, where $g^{(2)}$ is the second-order correlation function between two particles, as demonstrated via the Hanbury Brown and Twiss effect^{26–28}; however, the flux of correlated atom pairs is usually very low. Alternatively, the degree of correlation can be enhanced by making a source comprised largely of correlated atom pairs. In our experiment, we use *s*-wave collisions between ultracold atoms^{4,5}, which can yield a high degree of second-order (two-atom) correlation ($g^{(2)} \gg 2$). This set-up is analogous to correlated bi-photon sources in that pairs of correlated particles are created at the same point in space and time. Like SPDC, these sources can exhibit quantum correlations²⁹ and demonstrate Bell’s inequalities³⁰. In fact, the high correlations and small spatial correlation width of our source mean it would be well suited to such measurements, although in our ghost imaging experiment it is the strong correlations between atom pairs rather than any quantum entanglement that enables ghost imaging.

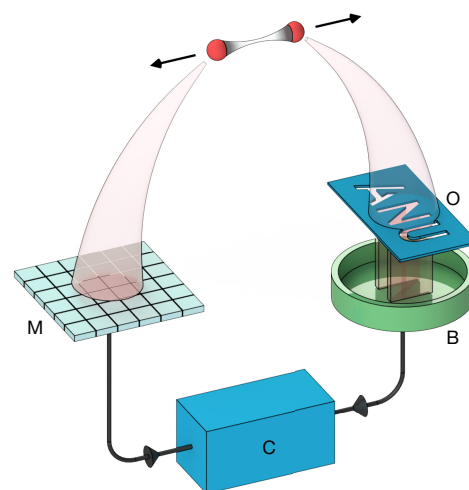


Figure 1 | Schematic of atomic ghost imaging. Correlated pairs of atoms created in a collision form two beams. One beam passes through the object to be imaged (‘O’) and the arrival times of the individual atoms are detected by a bucket detector (‘B’). The second beam never interacts with the object, but is detected with full temporal and spatial resolution by a multi-pixel detector (‘M’). A correlator (‘C’) then reconstructs the image of the object.

¹Research School of Physics and Engineering, Australian National University, Canberra, Australian Capital Territory 2601, Australia.

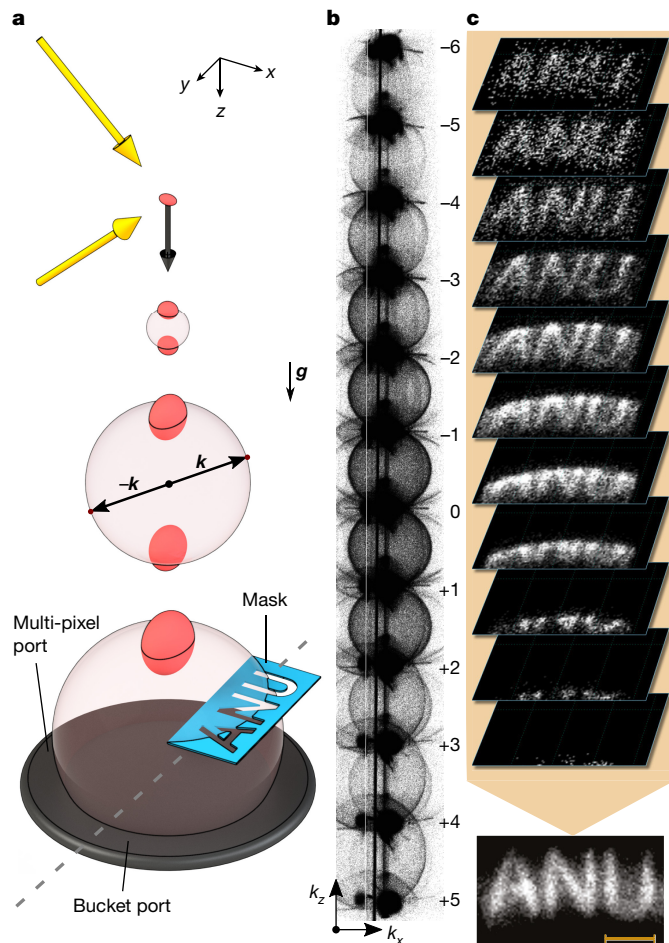


Figure 2 | Schematic of the experiment and resulting ghost image.

a, The experiment starts by using Kapitza-Dirac diffraction laser beams (yellow arrows) to split the trapped BEC cloud (dark red sphere) into different momentum states (for simplicity, only the first diffraction order is shown as the clouds evolve in time). Binary atomic collisions populate an *s*-wave scattering halo (pale red sphere) with correlated pairs of opposite momenta ($\pm k$), which then expand as they fall under gravity (g). In the momentum space associated with the centre-of-mass reference frame, the BECs are situated at opposite poles of the *s*-wave scattering sphere. Some of the halo atoms pass through a mask placed 10 mm above the bucket port of the single-atom detector, while their diametrically opposing counterparts (separated by the dashed grey line) are registered by the multi-pixel port. **b**, Experimental data from 2,000 individual experimental runs showing the 11 halos produced in the collision process, where the individual atom counts are reconstructed in three-dimensional momentum space. Kapitza-Dirac diffraction produces 12 diffraction orders $-6, \dots, +5$ along e_z . Collisions between each pair of adjacent orders result in 11 independent scattering halos. **c**, Individual ghost images from each of the halos from 68,835 experimental runs are combined to form the final image (bottom; scale bar, 5 mm). Because of the difference in absolute velocities for different diffraction orders, the halos that land first cover only a fraction of the image.

However, such a high degree of correlation usually requires a very low flux of atoms, which in turn would necessitate an extremely long data acquisition time. To overcome this problem, we developed a technique to enhance the number of correlated atom pairs, while maintaining a high degree of correlation and therefore signal-to-noise ratio. We use higher-order Kapitza-Dirac diffraction^{6–8} to produce multiple distinct sources of *s*-wave-scattered atoms in each experimental run. Using multiple sources enables more than a tenfold increase in the data acquisition rate, thus making the experiment feasible—the data we present here represent around three weeks of full-time data acquisition.

The experiments start with a BEC of helium atoms in the metastable (2^3S_1) state: $^4\text{He}^*$. This state enables single-atom detection with high efficiency because of the large internal energy of the atoms³. We magnetically trap a BEC of about one million $^4\text{He}^*$ atoms in the $m_J = +1$ sublevel with no discernible thermal fraction (see Methods for details). To produce an *s*-wave halo, we collide atoms in the BEC in two steps: (i) we out-couple nearly all atoms from the trap via a two-photon Raman transition to the magnetically insensitive sublevel $m_J = 0$ and (ii) we diffract the cloud into multiple momentum modes. This last step introduces a relative momentum difference between atoms in the different diffraction orders, which collide and thus generate a series of *s*-wave scattering halos via binary collisions (Methods). We use the same laser beams for both Raman and Kapitza-Dirac pulses, with the latter Fourier-broadened to ensure we are in the Kapitza-Dirac regime, which populates multiple diffraction orders. Both Raman and Kapitza-Dirac pulses propagate along the $(e_x \pm e_z)/\sqrt{2}$ directions as shown in Fig. 2a, where e_x , e_y and e_z are the unit vectors along the x , y and z directions, respectively. These propagation directions result in momentum transfer along the vertical z axis, with the momentum difference between any two adjacent diffracted orders $\hbar\Delta k = \pm\sqrt{2}\hbar k_0 e_z$, where $\hbar k_0$ is a single photon recoil, $k_0 = 2\pi/\lambda$ and $\lambda = 1,083.1979$ nm is the photon wavelength. In k -space associated with the centre-of-mass reference frame, each *s*-wave scattering halo comprises atoms on a sphere of radius $k_r = \Delta k/2 = k_0/\sqrt{2}$, which reflects energy and momentum conservation, as shown in Fig. 2a.

After the collision, the expanding halo falls approximately 850 mm under gravity onto a detector that consists of two multi-channel plates and a delay line, which records the arrival positions and time (x, y, t) of individual atoms. We then transform this three-dimensional information into the momenta (k_x, k_y, k_z) of individual atoms, which is plotted in Fig. 2b to show the 11 halos produced in the collision process. The dark clouds represent BECs in different diffraction orders ℓ with an *s*-wave halo situated between each of the corresponding orders ($\ell, \ell + 1$). The centre of each halo is on the momentum-transfer axis z . Also clearly visible are the portions of larger-diameter halos from non-adjacent orders ($\ell, \ell + 2$), ($\ell, \ell + 3$) and so on, as well as halos formed by single-photon spontaneous scattering from the diffraction laser. The halo populations follow that of different orders in the Kapitza-Dirac effect and, as a consequence, the average number of atoms per halo decreases from about 370 for $(\ell, \ell + 1) = (0, +1)$ to about 50 for $(+5, +6)$.

Ghost imaging is demonstrated by placing a thin metal mask 10 mm above the detector, which covers a portion of the detector's surface such that only a fraction of the *s*-wave halo (containing at most one atom

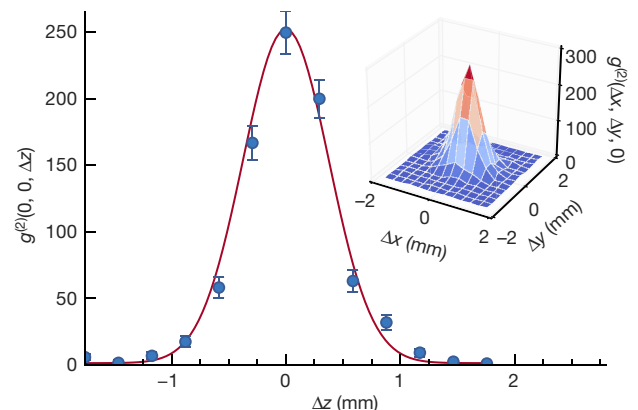


Figure 3 | Cross-correlation function. The main plot shows $g^{(2)}(0, 0, \Delta z)$ as a function of only the vertical coordinate Δz . The solid line is a Gaussian fit, which has an r.m.s. width of $\sigma_z = 0.37$ mm, corresponding to the correlation length. Error bars show the statistical error over the 54,473 experimental runs. The inset shows $g^{(2)}(\Delta x, \Delta y, 0)$ for the experimental data.

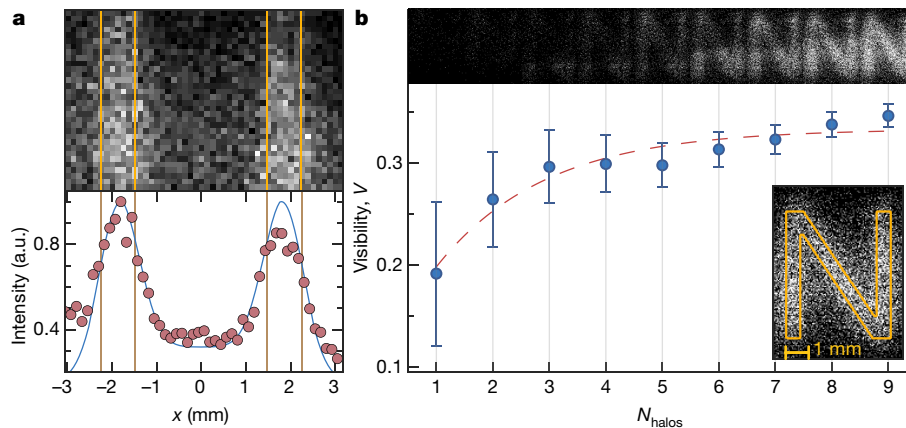


Figure 4 | Resolution and visibility of the ghost image. **a**, Ghost image (top) of the vertical bars of the ‘U’ (indicated by the yellow bars). The image is integrated vertically to yield an intensity (bottom, circles), which is fitted with the convolution with the PSF (bottom, solid line). This yields a width (representing the imaging resolution) of 0.40 mm. **b**, The visibility ($V = (I - B)/(I + B)$) is shown as a function of the number of halos (N_{halos})

from each correlated pair) passes through the mask. The rest of the atoms are detected directly, without any interaction with the object, as shown in Fig. 2a. The detector is artificially split into two regions: the bucket region, masked with the object we want to image, and the multi-pixel region, where the atoms are detected with full spatial and temporal resolution. On the bucket portion of the detector, only the time of arrival t is used for ghost image reconstruction, whereas for the multi-pixel detector we retain a full (x, y, t) set. The ghost image is then reconstructed in k -space using coincidence-counting between atoms in the multi-pixel port and the bucket port. Combining the independent images from halos in the different diffraction orders results in the ghost image shown at the bottom of Fig. 2c. The ghost image clearly resolves the object, which is a 15.5-mm-wide mask of the letters ‘ANU’ with line widths of approximately 0.75 mm.

Following ref. 4, we characterize the correlations between atoms with almost equal, but opposite, momenta $\mathbf{k}_1 = -\mathbf{k}_2 + \Delta\mathbf{k}$ (where $\Delta\mathbf{k} = (\Delta k_x, \Delta k_y, \Delta k_z)$) and the centre of each sphere has $\mathbf{k} = \mathbf{0}$ by constructing the two-particle cross-correlation function $\tilde{g}^{(2)}(\Delta\mathbf{k})$. Figure 3 shows the second-order spatial correlation function between the two atoms $g^{(2)}$ for the halo from diffraction orders $\ell = +3$ and $\ell = +4$. To facilitate analysis of the ghost image resolution, we convert $\tilde{g}^{(2)}(\Delta\mathbf{k})$ into spatial coordinates, denoted $g^{(2)}(\Delta\mathbf{r})$ (where $\Delta\mathbf{r} = (\Delta x, \Delta y, \Delta z)$; Methods). The key factor limiting the image resolution is the finite width of $g^{(2)}$ —the correlation length.

A smaller average number of counts per halo is beneficial, because it leads to a higher peak $g^{(2)}$ (ref. 4) and, consequently, a lower probability of registering false coincidence counts, which contribute to the ghost image background. Multi-order Kapitza–Dirac diffraction is therefore crucial for this type of experiment, because it allows a large number of relatively dilute s -wave halos to be populated and detected in a single experimental run, thereby substantially decreasing the required experimental run-time.

In the x - y plane of the detector, the absolute image resolution is limited by the real-space widths of $g^{(2)}(\Delta x, \Delta y, 0)$, whereas the spread along z contributes to background counts during image reconstruction. The $g^{(2)}(\Delta x, \Delta y, \Delta z)$ data in Fig. 3 are fitted with a three-dimensional Gaussian function $f(\Delta\mathbf{r}) = 1 + A \exp[-\Delta x^2/(2\sigma_x^2) - \Delta y^2/(2\sigma_y^2) - \Delta z^2/(2\sigma_z^2)]$, yielding root-mean-square (r.m.s.) widths of $\{\sigma_{x,y,z}\} = \{0.43, 0.39, 0.37\} \pm 0.01$ mm (errors given here and elsewhere represent the standard deviation), which are set by a combination of the spatial correlation length and the detector resolution.

Our ghost imaging resolution can also be estimated by analysing the image of a known shape, which we do in Fig. 4a for the bars of

that are accumulated to form the image. I is the mean intensity within the shape to be imaged (yellow ‘N’, as shown in the inset) and B is the mean intensity outside the ‘N’. Each plotted image is the result of accumulating reconstructed images from different s -wave halos. The dashed curve is a guide to the eye; error bars show the standard error of the mean across the image pixels.

the ‘U’, which are known to be 0.77 mm wide. Fitting the convolution of the object with a one-dimensional Gaussian point spread function (PSF) gives the r.m.s. width of the PSF as 0.40 ± 0.03 mm along the x direction (Methods). This value is in good agreement with the value of $\sigma_x = 0.43 \pm 0.01$ mm obtained from the analysis of the $g^{(2)}$ widths.

In Fig. 4b we plot the visibility $V = (I - B)/(I + B)$ of the ghost image²², where I is the average intensity of the ghost image within the mask region (yellow lines) and B is the average intensity outside this region. We show an increase in the visibility of the ghost image as we cumulatively add contributions from different diffraction orders. The optimal, final image utilizes nine halos and has a visibility of about 35%, in agreement with a simple analytical model for our experimental parameters (Methods).

In conclusion, by using higher-order Kapitza–Dirac diffraction, we are able to extend the technique of pair production via s -wave scattering and demonstrate more than a tenfold increase in the number of correlated pairs available for each single experimental run. Because we created a source of strongly correlated atom pairs, by analogy with SPDC we are able to generate very high second-order $g^{(2)}$ correlation values that greatly exceed the maximum thermal value of two, with maxima in our case reaching $g^{(2)}(0, 0, 0) \approx 250$. Using this source of correlated twin beams we performed ghost imaging with atoms, achieving good visibility and showing that the sub-millimetre resolution of the image is limited by the two-particle correlation function of the atomic momenta. Extending the techniques demonstrated in this experiment could enable fundamental tests of quantum mechanics using massive particles^{9,10}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 July; accepted 11 October 2016.

- Erkmen, B. I. & Shapiro, J. H. Ghost imaging: from quantum to classical to computational. *Adv. Opt. Photonics* **2**, 405–450 (2010).
- Shapiro, J. H. & Boyd, R. W. The physics of ghost imaging. *Quantum Inf. Process.* **11**, 949–993 (2012).
- Vassen, W. *et al.* Cold and trapped metastable noble gases. *Rev. Mod. Phys.* **84**, 175–210 (2012).
- Perrin, A. *et al.* Observation of atom pairs in spontaneous four-wave mixing of two colliding Bose–Einstein condensates. *Phys. Rev. Lett.* **99**, 150405 (2007).
- Jaskula, J.-C. *et al.* Sub-Poissonian number differences in four-wave mixing of matter waves. *Phys. Rev. Lett.* **105**, 190402 (2010).
- Kapitza, P. L. & Dirac, P. A. M. The reflection of electrons from standing light waves. *Math. Proc. Camb. Philos. Soc.* **29**, 297–300 (1933).
- Gould, P. L., Ruff, G. A. & Pritchard, D. E. Diffraction of atoms by light: the near-resonant Kapitza–Dirac effect. *Phys. Rev. Lett.* **56**, 827–830 (1986).

8. Ovchinnikov, Yu. B. *et al.* Diffraction of a released Bose–Einstein condensate by a pulsed standing light wave. *Phys. Rev. Lett.* **83**, 284–287 (1999).
9. Kofler, J. *et al.* Einstein–Podolsky–Rosen correlations from colliding Bose–Einstein condensates. *Phys. Rev. A* **86**, 032115 (2012).
10. Jack, B. *et al.* Holographic ghost imaging and the violation of a Bell inequality. *Phys. Rev. Lett.* **103**, 083602 (2009).
11. Pittman, T. B., Shih, Y. H., Strekalov, D. V. & Sergienko, A. V. Optical imaging by means of two-photon quantum entanglement. *Phys. Rev. A* **52**, R3429–R3432 (1995).
12. Strekalov, D. V., Sergienko, A. V., Klyshko, D. N. & Shih, Y. H. Observation of two-photon “ghost” interference and diffraction. *Phys. Rev. Lett.* **74**, 3600–3603 (1995).
13. Klyshko, D. N. Effect of focusing on photon correlation in parametric light scattering. *Sov. Phys. JETP* **67**, 1131–1135 (1988).
14. Belinskii, A. V. & Klyshko, D. N. Two-photon optics: diffraction, holography, and transformation of two-dimensional signals. *Sov. Phys. JETP* **78**, 259–262 (1994).
15. Hardy, N. D. & Shapiro, J. H. Computational ghost imaging versus imaging laser radar for three-dimensional imaging. *Phys. Rev. A* **87**, 023820 (2013).
16. Yuan, S., Yao, J., Liu, X., Zhou, X. & Li, Z. Cryptanalysis and security enhancement of optical cryptography based on computational ghost imaging. *Opt. Commun.* **365**, 180–185 (2016).
17. Yu, H. *et al.* Fourier-transform ghost imaging with hard X rays. *Phys. Rev. Lett.* **117**, 113901 (2016).
18. Pelliccia, D., Rack, A., Scheel, M., Cantelli, V. & Paganin, D. M. Experimental X-ray ghost imaging. *Phys. Rev. Lett.* **117**, 113902 (2016).
19. Sun, B. *et al.* 3D computational imaging with single-pixel detectors. *Science* **340**, 844–847 (2013).
20. Ryzkowski, P., Barbier, M., Friberg, A. T., Dudley, J. M. & Genty, G. Ghost imaging in the time domain. *Nat. Photon.* **10**, 167–170 (2016).
21. Brida, G., Genovese, M. & Ruo Berchera, I. Experimental realization of sub-shot-noise quantum imaging. *Nat. Photon.* **4**, 227–230 (2010).
22. Morris, P. A., Aspden, R., Bell, J. E. C., Boyd, R. W. & Padgett, M. J. Imaging with a small number of photons. *Nat. Commun.* **6**, 5913 (2015).
23. Bennink, R. S., Bentley, S. J. & Boyd, R. W. “Two-photon” coincidence imaging with a classical source. *Phys. Rev. Lett.* **89**, 113601 (2002).
24. Bennink, R. S., Bentley, S. J., Boyd, R. W. & Howell, J. C. Quantum and classical coincidence imaging. *Phys. Rev. Lett.* **92**, 033601 (2004).
25. Ferri, F. *et al.* High-resolution ghost image and ghost diffraction experiments with thermal light. *Phys. Rev. Lett.* **94**, 183602 (2005).
26. Yasuda, M. & Shimizu, F. Observation of two-atom correlation of an ultracold neon atomic beam. *Phys. Rev. Lett.* **77**, 3090–3093 (1996).
27. Jelte, T. *et al.* Comparison of the Hanbury Brown–Twiss effect for bosons and fermions. *Nature* **445**, 402–405 (2007).
28. Hodgman, S. S., Dall, R. G., Manning, A. G., Baldwin, K. G. H. & Truscott, A. G. Direct measurement of long-range third-order coherence in Bose–Einstein condensates. *Science* **331**, 1046–1049 (2011).
29. Lewis-Swan, R. J. & Kheruntsyan, K. V. Proposal for demonstrating the Hong–Ou–Mandel effect with matter waves. *Nat. Commun.* **5**, 3752 (2014).
30. Lewis-Swan, R. J. & Kheruntsyan, K. V. Proposal for a motional-state Bell inequality test with ultracold atoms. *Phys. Rev. A* **91**, 052114 (2015).

Acknowledgements A.G.T. acknowledges the support of the Australian Research Council (ARC) through the Future Fellowship grant FT100100468 and the Discovery grant DP120101390. S.S.H. acknowledges the support of the ARC through the DECRA Fellowship DE150100315. We thank A. T. Friberg for discussions.

Author Contributions R.I.K., R.G.D. and A.G.T. conceived the experiment. R.I.K. performed the experiment and collected the data. All authors contributed to the conceptual formulation of the physics, the interpretation of the data and the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.G.T. (andrew.truscott@anu.edu.au).

Reviewer Information *Nature* thanks M. A. Kasevich and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Experimental apparatus. The He* BEC is initially trapped in a bi-planar quadrupole Ioffe-configuration magnetic trap³¹, with harmonic frequencies of $\{\omega_x, \omega_y, \omega_z\}/(2\pi) \approx \{15, 25, 25\}$ Hz and a bias field of $B_0 = 1.31 \pm 0.01$ G along the x axis. The 80-mm-diameter multi-channel plate and delay-line detector are located about 850 mm below the trap centre (corresponding to a fall time of about 416 ms), with a spatial resolution in x and y of approximately 120 μm , a temporal resolution along z of approximately 2 ns and a quantum efficiency of around 10%.

Diffraction. Similarly to previous work³², we use the same laser beams for both Raman and Kapitza–Dirac pulses, changing only the relative frequency detuning of the waveforms, which is set by the bias B_0 and the geometric angle between the beams (90°). The laser is blue-detuned by 2 GHz from the $2^3\text{S}_1 \rightarrow 2^3\text{P}_0$ transition. The Raman transition in the beginning of the sequence also results in the cloud acquiring a momentum change of $\sqrt{2}\hbar k_0 e_z$, which sets the centre-of-mass of the cloud in downward motion. The duration of the out-coupling Raman π pulse is $t_\pi = 1.8 \mu\text{s}$, with a transfer efficiency of about 95%. This pulse is immediately followed by a 1.4- μs Kapitza–Dirac pulse. The maximum intensity of each laser beam is approximately 450 mW mm^{-2} for the Raman pulse and 30 mW mm^{-2} for the Kapitza–Dirac pulse. Each of the pulse sequences is modulated with an overall Gaussian envelope to control Fourier broadening. Thus, the broadening of the Raman pulse was optimized to maximize the transfer to the magnetically insensitive sublevel $m_J = 0$, and the Kapitza–Dirac pulse was set to maximize the number of diffraction orders generated (L).

In our experiment we achieved $L = 12$ orders, which consequently produced $L - 1 = 11$ s -wave halos between the adjacent orders, as pictured in Fig. 2b. In principle, by increasing the laser power of the diffraction beams, we could raise L even further, generating more halos with lower density and therefore possessing higher correlations. However, it is optimal to have $L = 12$ for the ghost imaging configuration in our experiment because of the relatively high absolute velocity of the $^4\text{He}^*$ atoms in the higher diffraction orders: for $\ell = \pm 6$, $v_z = \pm 78 \text{ cm s}^{-1}$. There are two problems that arise for $|\ell| > 6$ in our set-up: (i) fast travelling, downwards-kicked halos do not expand enough to cover the object mask before falling onto the detector (Fig. 2c), and (ii) upwards-kicked halos will be lost through hitting the top of the vacuum chamber.

We also observe larger-diameter halos from the non-adjacent orders ($\ell, \ell + 2$), which are reconstructed only partially because they expand beyond the detector size. In principle, $g^{(2)}$ peaks should be higher for these larger halos, because the scattering modes overlap less⁴. However, the population of these halos is too low to make them usable here.

The non-zero halo thickness δk (r.m.s.) is due to uncertainty-limited broadening from the finite size of the condensates, as well as mean-field interactions in the BEC during expansion of the halo. In our experiments, we observe $\delta k/k_r = 0.034 \pm 0.003$.

Correlation function. For independent s -wave halos originating from different diffraction orders (Fig. 2b), we construct a second-order normalized cross-correlation function in k -space:

$$\tilde{g}^{(2)}(\Delta\mathbf{k}) = \frac{\int G^{(2)}(\mathbf{k}, -\mathbf{k} + \Delta\mathbf{k}) d^3\mathbf{k}}{\int \langle n(\mathbf{k}) \rangle \langle n(-\mathbf{k} + \Delta\mathbf{k}) \rangle d^3\mathbf{k}} \quad (1)$$

where $G^{(2)}(\mathbf{k}, -\mathbf{k} + \Delta\mathbf{k})$ is the un-normalized two-particle cross-correlation function. Because of the spherical symmetry of the s -wave halo in momentum space, it is convenient to perform the analysis with atomic momenta. However, to allow a spatial resolution to be extracted, the final results are converted to spatial units as $g^{(2)}(\Delta\mathbf{r})$. Because we know the time-of-flight to the detector T_f , we can change the coordinate representations of the detected atom position, from velocity (momentum) space to position space or position and time: (v_x, v_y, v_z) or $(k_x, k_y, k_z) \rightarrow (x, y, z) \rightarrow (x, y, t)$. For the spatial coordinate representation for $g^{(2)}(\Delta\mathbf{r})$ in Fig. 3, we convert atom velocities $\Delta\mathbf{v} = \hbar\Delta\mathbf{k}/m$, where m is the mass of a $^4\text{He}^*$ atom, to spatial coordinates in the x - y plane of the detector: $\Delta\mathbf{r} = T_f\Delta\mathbf{v}$. Similar to previous work³³, the normalization in Equation (1) is performed by calculating $G^{(2)}(\mathbf{k}, -\mathbf{k} + \Delta\mathbf{k})$ across different experimental runs, because the atoms will be uncorrelated between runs.

We find that the r.m.s. widths of $g^{(2)}$ are approximately constant across all diffraction orders. Note that the correlation lengths we measure are actually a convolution of the real correlation length with the detector resolution (approximately 120 μm).

The error estimate in the numerical evaluation of $g^{(2)}$ can be expressed as a standard deviation of the atom count frequencies (which represent the numerator of Equation (1) normalized to the same denominator). The error bars in Fig. 3 represent a 2σ confidence interval. The data used to produce Fig. 3 were taken from 54,473 experimental runs under identical experimental conditions to those used

to produce the ghost image, but without the object mask in front of the detector because the mask blocks a large portion of the halos.

Ghost image. The target to be imaged was made out of a laser-cut 0.4-mm-thick stainless steel sheet. A microscope image of the object used for the ghost imaging is shown in Extended Data Fig. 1.

The ghost image is reconstructed in k -space using coincidence-counting between the atoms detected in the multi-pixel port n_M and the bucket port n_B :

$$\tilde{C}(k_x, k_y) = n_M(k_x, k_y, k_z) \otimes n_B(-k_z + \Delta k_z) \quad (2)$$

The correlator in Equation (2) is calculated in momentum space and accepts two counts from n_M and n_B as a valid coincidence if they occur within the correlation length along the z (time) direction, which in spatial coordinates corresponds to $|\Delta z| \leq \sigma_z$. The correlation length σ_z is given by the $g^{(2)}$ width of 0.37 mm (see Fig. 3). If the two counts are determined to be correlated, then the coordinates (k_x, k_y) on the multi-pixel detector are recorded and added to the image. Next, analogously to $\tilde{g}^{(2)}$, we convert the image $\tilde{C}(k_x, k_y)$ to spatial coordinates $C(x, y)$ in the detector plane $z = 0$. This image is shown in Fig. 2c.

Image resolution and visibility. We estimate the resolution of our ghost imaging configuration by taking the ghost image of the vertical bars from the ‘U’ of the ‘ANU’ mask (Extended Data Fig. 1), and fitting it with the convolution $Q(x) = b * h$ of the real object shape $b(x)$ and the Gaussian PSF $h(x) = A + C \exp[-x^2/(2\sigma_x^2)]$. Therefore, using the r.m.s. width σ_x of the PSF as one of three free parameters of the fit, we can compare the resolution of the ghost image along the x direction with the correlation length along that axis. As stated in the main text, we find $\sigma_x = 0.40 \pm 0.03$ mm, which is in good agreement with the r.m.s. width of $\sigma_x = 0.43 \pm 0.01$ mm extracted from $g^{(2)}$.

We characterize the visibility of the ghost image by comparing the average intensity I within the mask region to the average intensity B outside the mask:

$$V = \frac{I - B}{I + B} \quad (3)$$

where I is calculated within the region that matches the object mask. This region is determined by manually overlapping an ‘N’-shaped region with the exact dimensions of the mask on top of the ghost image.

The visibility will be reduced from the ideal value of one as a result of three main factors. First, any atom pairs in the halo that are uncorrelated, but arrive within the same time window, will result in a false count, which will reduce the visibility; however, given the high degree of correlation present in our halos (as evidenced by the large $g^{(2)}$ peaks), this will be a small contribution. Second, there are stray background counts due to particles not generated by s -wave scattering (about 2.5 counts per halo in the region of interest) and to the dark count rate of the multi-channel plate (about 0.2 counts per halo), which can lead to false correlations. Finally, the major limiting factor to our visibility is the spatial resolution of our imaging procedure ($\sigma_{x,y} \approx 400 \mu\text{m}$), which is related to the detector resolution and the finite spatial correlation length, and results in blurring around the edges of the image. This blurring has a particularly strong effect on the visibility because the smallest features of the mask are within a factor of two larger than the spatial resolution.

The effect of spatial resolution on visibility is illustrated by an estimate of V in the x dimension, where the mask feature size d_x is around twice the correlation length: $d_x \approx 2\sigma_x$. The ratio I/B for this regime can be approximated by exploiting the Gaussian shape of the convolving PSF $h(x)$ (width σ_x) as follows. If calculated over approximately equal areas of the image, both I and B are proportional to $\int h(x) dx$ (because the integral of the convolution here is approximately equal to the product of the integrals). I is integrated over the object region $|x| \leq \sigma_x$, whereas B is calculated outside the object, $|x| > \sigma_x$. Thus, we obtain $I/B = 0.68/0.32$, yielding $V \approx 0.36$, which is in good agreement with the visibility of $V \approx 0.35$ for the ‘N’ in Fig. 4.

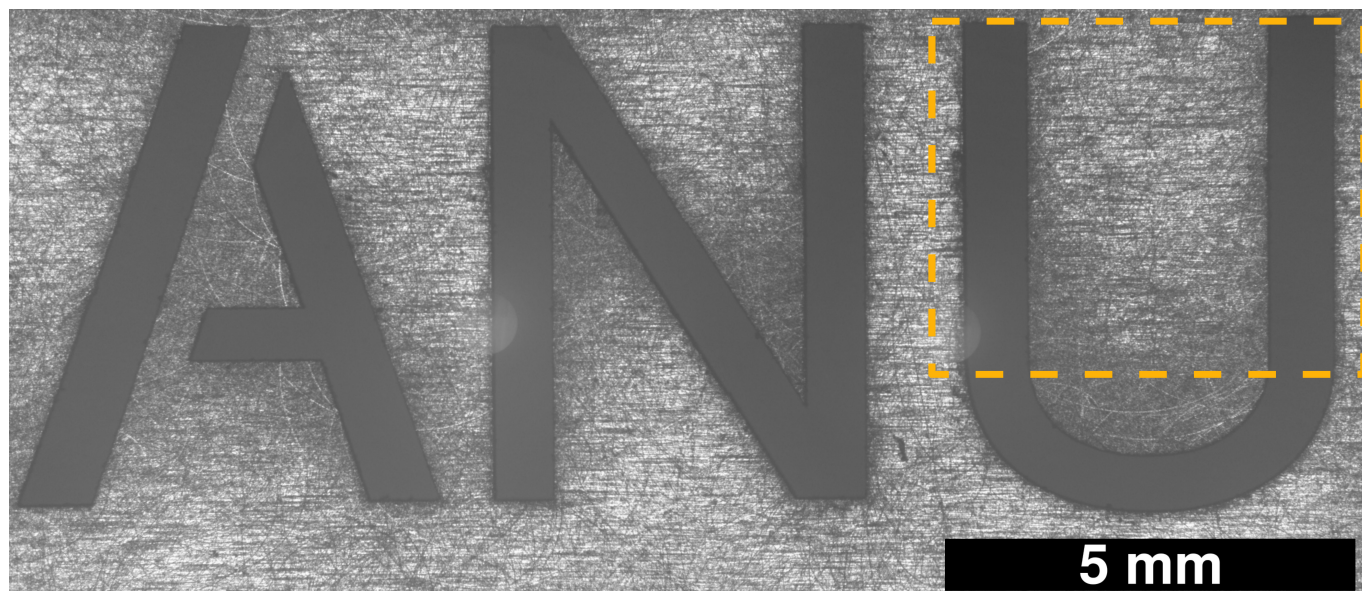
Because the ghost imaging resolution $\sigma_{x,y}$ is a combination of the detector resolution and the spatial correlation length of the halo, reducing either would yield better visibility. In our case these contributions are roughly equal, so making either of them negligible (that is, in the limit of either perfect detector resolution or zero spatial correlation length) would improve our resolution by a factor of about two, which in the above integration would in principle increase the visibility to $V \approx 0.9$. However, in reality we would not achieve such high visibility, because false coincidences due to uncorrelated atoms or dark counts of the detector limit our ultimate visibility to $V \approx 0.8$, even for perfect spatial resolution and a negligible correlation length.

Extended Data Fig. 2 shows the ghost image visibility for images from each individual halo as a function of the average atom number $\langle N_a \rangle$ in that halo. The inset shows the corresponding object (the letter ‘N’) and its ghost image for the halo with the highest number of counts. Because of the different centre-of-mass

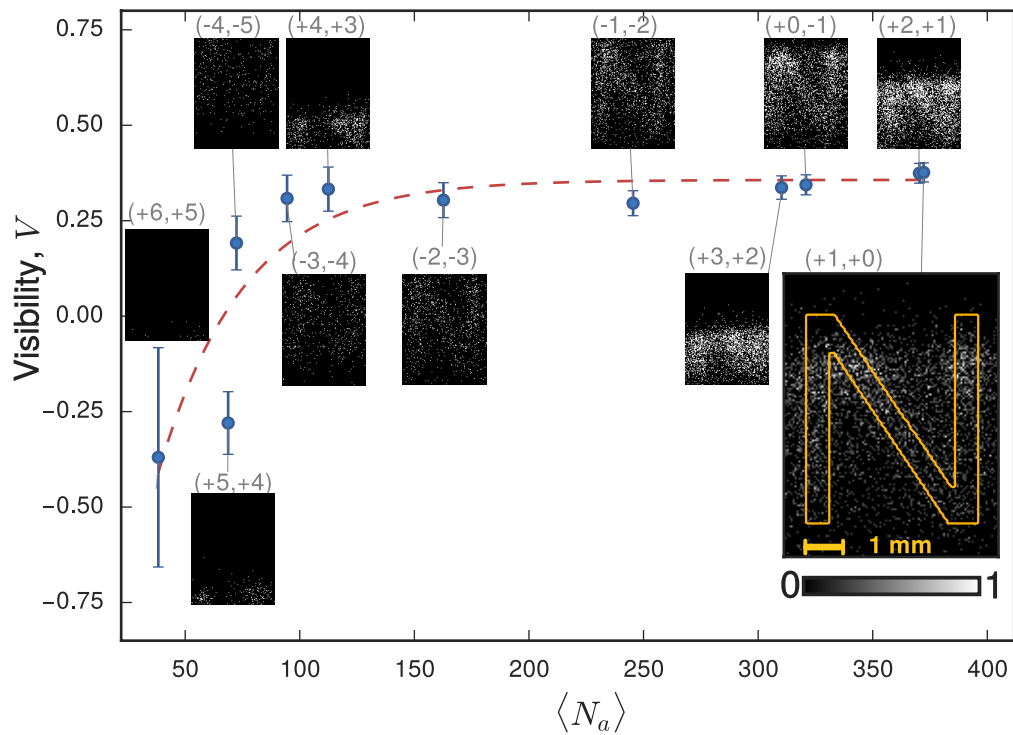
velocities, the halos with the smallest time-of-flight only partially cover the mask. The first two halos to arrive onto the detector, (+6, +5) and (+5, +4), have a small halo population (N_h) and minimal overlap with the mask, which results in a weak ghost image signal relative to the background and hence $V < 0$. Therefore, these halos were not used to produce the cumulative ghost image shown in Fig. 4. Equation (3) gives high visibility values (many counts within the object region and a small background) even if the halos do not cover the object completely and thus do not generate a full representation of the entire image. However, we used this visibility definition to be consistent with that used in the photon ghost imaging literature²².

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

31. Dall, R. G. & Truscott, A. G. Bose–Einstein condensation of metastable helium in a bi-planar quadrupole Ioffe configuration trap. *Opt. Commun.* **270**, 255–261 (2007).
32. Manning, A. G., Khakimov, R. I., Dall, R. G. & Truscott, A. G. Wheeler's delayed-choice gedanken experiment with a single atom. *Nat. Phys.* **11**, 539–542 (2015).
33. Dall, R. G. *et al.* Ideal n -body correlations with massive particles. *Nat. Phys.* **9**, 341–344 (2013).



Extended Data Figure 1 | The object. Microscope image of the mask used to create the ghost image. The region indicated by the dashed line forms the vertical bars shown in Fig. 4a, which was used to determine the ghost imaging resolution.



Extended Data Figure 2 | Ghost image visibility. Visibilities (dots) for images (insets) reconstructed from each individual halo with different average numbers of atoms $\langle N_a \rangle$. Diffraction orders producing the halos are labelled as $(\ell + 1, \ell)$. The dashed curve is a guide to the eye. Error bars represent the standard error of the mean associated with the variances of the pixel values contributing to I and B .

Quantifying global soil carbon losses in response to warming

T. W. Crowther^{1,2}, K. E. O. Todd-Brown³, C. W. Rowe², W. R. Wieder^{4,5}, J. C. Carey⁶, M. B. Machmuller⁷, B. L. Snoek^{1,8}, S. Fang^{9,10}, G. Zhou⁹, S. D. Allison^{11,12}, J. M. Blair¹³, S. D. Bridgman¹⁴, A. J. Burton¹⁵, Y. Carrillo¹⁶, P. B. Reich^{16,17}, J. S. Clark¹⁸, A. T. Classen^{19,20}, F. A. Dijkstra²¹, B. Elberling²², B. A. Emmett²³, M. Estiarte^{24,25}, S. D. Frey²⁶, J. Guo²⁷, J. Harte²⁸, L. Jiang²⁹, B. R. Johnson³⁰, G. Kröel-Dulay³¹, K. S. Larsen³², H. Laudon³³, J. M. Lavelle^{7,34}, Y. Luo^{29,35}, M. Lupascu³⁶, L. N. Ma³⁷, S. Marhan³⁸, A. Michelsen^{22,39}, J. Mohan⁴⁰, S. Niu⁴¹, E. Pendall¹⁶, J. Peñuelas^{24,25}, L. Pfeifer-Meister¹⁴, C. Poll³⁸, S. Reinsch²³, L. L. Reynolds¹⁴, I. K. Schmidt³², S. Sistla⁴², N. W. Sokol³, P. H. Templer⁴³, K. K. Treseder¹², J. M. Welker⁴⁴ & M. A. Bradford^{1,2}

The majority of the Earth's terrestrial carbon is stored in the soil. If anthropogenic warming stimulates the loss of this carbon to the atmosphere, it could drive further planetary warming^{1–4}. Despite evidence that warming enhances carbon fluxes to and from the soil^{5,6}, the net global balance between these responses remains uncertain. Here we present a comprehensive analysis of warming-induced changes in soil carbon stocks by assembling data from 49 field experiments located across North America, Europe and Asia. We find that the effects of warming are contingent on the size of the initial soil carbon stock, with considerable losses occurring in high-latitude areas. By extrapolating this empirical relationship to the global scale, we provide estimates of soil carbon sensitivity to warming that may help to constrain Earth system model projections. Our empirical relationship suggests that global soil carbon stocks in the upper soil horizons will fall by 30 ± 30 petagrams of carbon to 203 ± 161 petagrams of carbon under one degree of warming, depending on the rate at which the effects of warming are realized. Under the conservative assumption that the response of soil carbon to warming occurs within a year, a business-as-usual climate scenario would drive the loss of 55 ± 50 petagrams of carbon from the upper soil horizons by 2050. This value is around 12–17 per cent of the expected anthropogenic emissions over this period^{7,8}. Despite the considerable uncertainty in our estimates, the direction of the global soil carbon response is consistent across all scenarios. This provides strong empirical support for the idea that rising temperatures will stimulate the net loss of soil carbon to the

atmosphere, driving a positive land carbon–climate feedback that could accelerate climate change.

The exchange of carbon (C) between the soil and atmosphere represent a prominent control on atmospheric C concentrations and the climate^{1,6,9}. These processes are driven by the organisms (plants, microbes and animals) that live in the soil, the activity of which could be accelerated by anthropogenic warming¹⁰. If warming stimulates the loss of C into the atmosphere, it could drive a land C–climate feedback that could accelerate climate change. Yet despite considerable scientific attention in recent decades, there remains no consensus on the direction or magnitude of warming-induced changes in soil C^{11,12}. There is growing confidence that warming generally enhances fluxes to and from the soil^{8,12}, but the net global balance between these responses remains uncertain and direct estimates of soil C stocks are limited to single-site experiments that generally reveal no detectable effects^{5,13–15}.

Given the paucity of direct measurements of the responses of soil C stocks to warming, Earth system models (ESMs) must rely heavily on the short-term temperature responses of soil respiration (Q_{10}) to infer long-term changes in global C stocks. Without empirical observations that capture longer-term C dynamics, we are limited in our ability to evaluate model performance or to constrain the uncertainty in model projections¹⁶. As such, the land C–climate feedback remains one of the largest sources of uncertainty in current ESMs^{12,14,17}, restricting our capacity to develop C emissions targets that are compatible with specific climate change scenarios. Direct field measurements

¹Netherlands Institute of Ecology, Droevendaalsesteeg 10, 6708 PB Wageningen, The Netherlands. ²Yale School of Forestry & Environmental Studies, Yale University, 370 Prospect Street, New Haven, Connecticut 06511, USA. ³Pacific Northwest National Laboratory, Richland, Washington, Washington 99354, USA. ⁴Climate & Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado 80307, USA. ⁵Institute of Arctic & Alpine Research, University of Colorado, Boulder, Colorado 80303, USA. ⁶Marine Biological Laboratory, 7 MBL Street, Woods Hole, Massachusetts 02543, USA. ⁷Natural Resource Ecology Laboratory, 1499 Campus Delivery, Colorado State University, Fort Collins, Colorado 80523-1499, USA. ⁸Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. ⁹Chinese Academy of Meteorological Sciences, No. 46 Zhongguancun South Street, Beijing 100081, China. ¹⁰Collaborative Innovation Center on Forecast Meteorological Disaster Warning & Assessment, Nanjing University of Information Science & Technology, Nanjing 210044, China. ¹¹Department of Earth System Science, University of California Irvine, Irvine, California 92697, USA. ¹²Department of Ecology & Evolutionary Biology, University of California Irvine, California 92697, USA. ¹³Division of Biology, Kansas State University, Manhattan, Kansas 66506, USA. ¹⁴Institute of Ecology & Evolution, University of Oregon, Eugene, Oregon 97403, USA. ¹⁵School of Forest Resources & Environmental Science, Michigan Technological University, Houghton, Michigan 49931, USA. ¹⁶Hawkesbury Institute for the Environment, Western Sydney University, Penrith, 2570 New South Wales, Australia. ¹⁷Department of Forest Resources, University of Minnesota, St. Paul, Minnesota 55108, USA. ¹⁸Nicholas School of the Environment, Duke University, Durham, North Carolina 27708, USA. ¹⁹The Center for Macroecology, Evolution, and Climate, The Natural History Museum of Denmark, University of Copenhagen, Universitetsparken, 15, 2100, København Ø, Denmark. ²⁰Department of Ecology & Evolutionary Biology, University of Tennessee, 569 Dabney Hall, 1416 Circle Drive, Knoxville, Tennessee 37996, USA. ²¹Centre for Carbon, Water & Food, The University of Sydney, Camden, 2570 New South Wales, Australia. ²²Center for Permafrost (CENPERM), Department of Geosciences and Natural Resource Management, University of Copenhagen, Øster Voldgade 10, 1350 Copenhagen K, Denmark. ²³Centre for Ecology and Hydrology, Environment Centre Wales, Deiniol Road, Bangor LL57 2UW, UK. ²⁴CSIC, Global Ecology Unit CREAF-CSIC, Cerdanyola del Vallès, 08193 Catalonia, Spain. ²⁵CREAF, Cerdanyola del Vallès, 08193 Catalonia, Spain. ²⁶Department of Natural Resources & the Environment, University of New Hampshire, Durham, New Hampshire 03824, USA. ²⁷Key Laboratory of Vegetation Ecology, Ministry of Education, Northeast Normal University, Changchun 130024, Jilin Province, China. ²⁸Energy & Resources Group, University of California at Berkeley, Berkeley, California 94720, USA. ²⁹Department of Microbiology & Plant Biology, University of Oklahoma, Norman, Oklahoma 73019, USA. ³⁰Department of Landscape Architecture, University of Oregon, Eugene, Oregon 97403, USA. ³¹Institute of Ecology & Botany, Magyar Tudományok Akadémia Centre for Ecological Research, 2–4 Alkotmány Utcakereso, Vacratot 2163, Hungary. ³²Department of Geosciences & Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark. ³³Department of Forest Ecology & Management, Swedish University of Agricultural Sciences, 90183 Umeå, Sweden. ³⁴Faculty of Life Sciences, University of Manchester, Dover Street, Manchester M13 9PT, UK. ³⁵Center for Earth System Science, Tsinghua University, Beijing 100084, China. ³⁶Department of Geography, National University of Singapore, 1 Arts Link, 117570 Singapore, Singapore. ³⁷State Key Laboratory of Vegetation & Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. ³⁸Institute of Soil Science & Land Evaluation, University of Hohenheim, 70593 Stuttgart, Germany. ³⁹Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark. ⁴⁰Odum School of Ecology, University of Georgia, Athens, Georgia 30601, USA. ⁴¹Key Laboratory of Ecosystem Network Observation & Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China. ⁴²School of Natural Science, Hampshire College, 893 West Street, Amherst, Massachusetts 01002, USA. ⁴³Department of Biology, Boston University, Boston, Massachusetts 02215, USA. ⁴⁴Department of Biological Sciences, University of Alaska, Anchorage, Anchorage, Alaska 99508, USA.

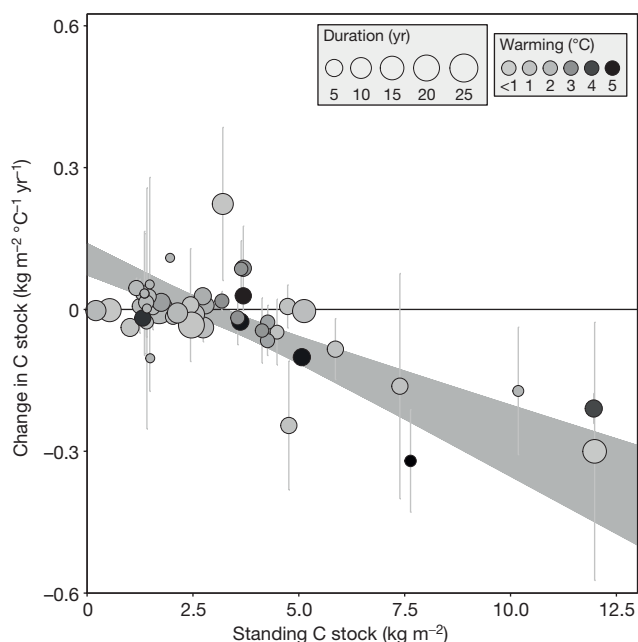


Figure 1 | The effect of warming on soil C losses depends on the initial standing soil C stock. The interaction between warming (degree-years) and standing C stocks is a primary determinant of the final warmed soil C stocks in the top 10 cm of soil (estimated using a mixed effects model; $n = 229$; see Supplementary Information). Each point represents the difference (mean \pm standard error) between soil C stocks in warmed and ambient plots within an individual experiment. The size of each point represents the length of the individual study and the colour indicates the amount of warming. The shaded area represents the bootstrapped 95% confidence interval ($R^2 = 0.49$; see Supplementary Information for details).

of warming-induced changes in soil C stocks are urgently needed to increase confidence in future climate projections¹⁶.

We took advantage of the growing number of climate change experiments around the world to compile a global database of soil C stock responses to warming. Soil samples were collected from replicate plots in 49 climate change experiments conducted across six biomes, ranging from arctic permafrost to dry Mediterranean forests (Extended Data Fig. 1). We compared soil C stocks across warmed (treatment) and ambient (control) plots to explore the effects of temperature across sites. The measured differences in the soil C stocks represent the net result of long-term changes in soil C inputs (plant production) and outputs (respiration) in response to warming. By linking these soil C responses to climatic and soil characteristics, we are able to generate a spatial understanding of the temperature sensitivity of soil C stocks at a global scale. To standardize collection protocols and account for the considerable variability in soil horizon depths, we focus on C stocks in the top 10 cm of soil. At a global scale, this upper soil horizon contains the greatest proportion of biologically active soil C⁹.

The effects of warming on soil C stocks were variable, with positive, negative and neutral impacts observed across sites (Fig. 1). However, the direction and magnitude of these warming-induced changes were predictable (Fig. 2) as they are contingent on the size of the standing soil C stocks and the extent and duration of warming. The interaction between control C stocks and degree-years (the standardized metric used to represent the multiplicative product of the extent (in °C) and duration (in years) of warming) was a strong explanatory variable when predicting warmed C stocks (additive model Akaike information criterion (AIC) = 383 versus multiplicative model AIC = 381; see Supplementary Information and Equation (1)). Specifically, the effects of warming were negligible in areas with small initial C stocks, but losses occurred beyond a threshold of 2–5 kg C m⁻² and were considerable in soils with ≥ 7 kg C m⁻²

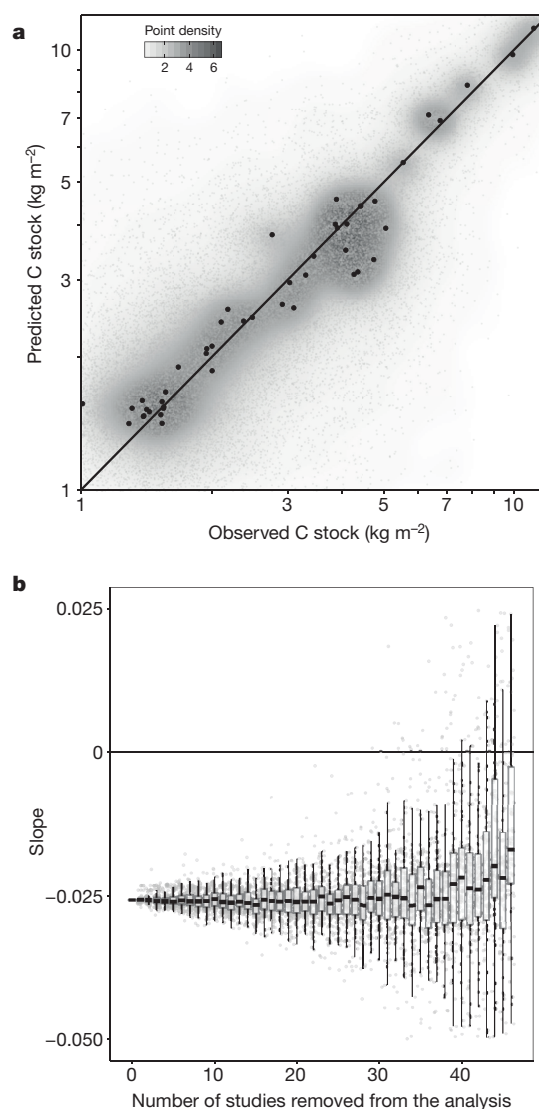


Figure 2 | Validation plots highlighting the predictive strength of the statistical model. **a**, Predicted versus observed soil C stock values in warmed treatment plots (estimated using statistical Equation (1), $R^2 = 0.95$; this high value is driven by the correlation between C values in the control and warmed plots). The black points represent the mean values for each study, and the shaded area represents the density of 1,000 simulated points randomly selected from within the normal distribution for each study. The 1:1 line is included to highlight perfect correspondence between the predicted and observed points and distributions. **b**, Bootstrapped estimates of the model (Equation (2)) slope values for different sample sizes. Studies were removed at random, the slope coefficient was calculated and this was repeated 1,000 times. Each point represents a bootstrapped estimate of slope for the model that included any given number of studies, and we include the interquartile range and median slope estimates for each number. The average slope value remains unchanged until >38 studies have been removed from the initial analysis (with 49 studies), highlighting that the relationship we present is not disproportionately influenced by the effects of warming in any specific study or site.

(Fig. 1). No other environmental characteristics (mean annual temperature, precipitation, soil texture or pH) significantly ($P > 0.1$) influenced the responses of soil C stocks to warming in our statistical models (additive environmental with degree-year model AIC = 388; see Supplementary Information).

The dominant role of standing C stocks in governing the magnitude of warming-induced soil C losses is in line with both empirical and theoretical expectations^{17–19}. The thawing of permafrost soils, where

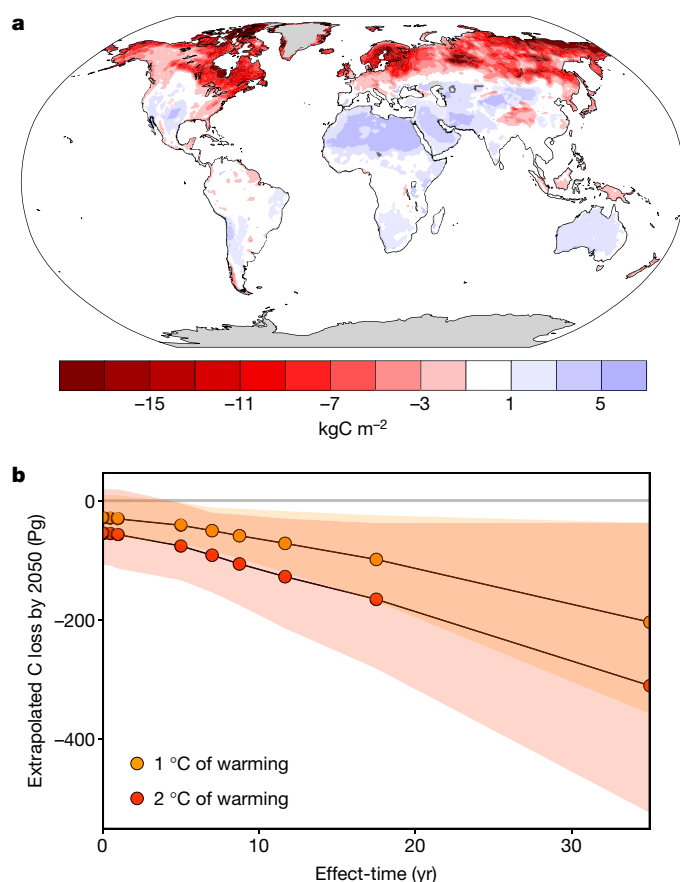


Figure 3 | Spatial extrapolation of the temperature vulnerability of soil C stocks. **a**, Map of predicted changes in soil C stocks per pixel by 2050 under the 'no acclimatization' scenario. This map was generated by extrapolating Equation (2) using spatially explicit estimates of soil C stocks¹⁹ and soil surface temperature change²² to reveal the spatial variation in projected changes in surface soil C stocks (0–15 cm depth) expected under a 1 °C rise in global average soil surface temperature. Note that Equation (2) reflects the maximum effect-time scenario, which generates the largest possible estimates of soil C change. This map also predicts C gains in tropical/desert regions that contain almost no soil C at present, but our lack of data in these mid-latitude regions means that we have low confidence in these effects. **b**, Total reductions in the global C pool under 1 °C and 2 °C global average soil surface warming by 2050, as expected under a full range of different soil C effect-time scenarios (*x* axis). Note that effect-time refers to the rate at which the full soil C response to warming is realized. Shaded areas indicate the 95% confidence intervals around the average C losses (dots) for each scenario. The rapid effect-time scenarios (for example, one week to one year) result in lower total soil C losses than the maximum effect-time scenario, but all simulations reveal considerable global losses of soil C under warming over the next 35 years.

limited C decomposition has led to the accumulation of large C stocks, will undoubtedly contribute to this phenomenon^{20,21}. However, our analysis also revealed considerable soil C losses in several non-permafrost regions, suggesting that additional mechanisms may contribute to the vulnerability of large soil C stocks. Presumably, the vulnerability of soils that contain large C stocks stems from the high temperature sensitivity of C decomposition and biogeochemical restrictions on the processes driving soil C inputs. In ecosystems with low initial soil C stocks, minor losses that result from accelerated decomposition under warming may be offset by concurrent increases in plant growth and soil C stabilization^{13,22}. In contrast, in areas with larger standing soil C stocks, accelerated decomposition outpaces potential C accumulation from enhanced plant growth, driving considerable C losses to the atmosphere.

By combining our measured soil C responses with spatially explicit estimates of standing C stocks²⁰ and soil surface temperature change²³, we reveal the global patterns in the vulnerability of soil C stocks (Fig. 3). Given that high-latitude regions have the largest standing soil C stocks²⁰ and the fastest expected rates of warming^{18,23}, our results suggest that the overwhelming majority of warming-induced soil C losses are likely to occur in Arctic and subarctic regions (Fig. 3). These high-latitude C losses considerably outweigh any minor changes expected in mid- and lower-latitude regions, providing further support for the idea of Arctic amplification of climate change feedbacks¹⁸ (Fig. 3). These warming-induced soil C losses need to be considered in light of future changes in moisture stress and vegetation growth, which are also likely to increase disproportionately in high-latitude areas¹⁸. Notably, the spatial distribution of soil C changes from our extrapolation contradicts projections from the CMIP5 archive of ESMs²⁴, which show increases in soil C at high latitudes—presumably due to the increases in plant productivity²⁵. The warming-induced losses of soil C that we observe have the potential to offset these vegetation responses, emphasizing the importance of representing soil C vulnerability in the process-based models that are used in climate change projections.

We extrapolated this relationship over the next 35 years to indicate how global soil C stocks might respond by 2050. The simple extrapolation of our empirical relationship suggests that 1 °C of warming over 35 years would drive the loss of 203 ± 161 PgC from the upper soil horizon (Fig. 3). However, this approach implicitly assumes that the effects of a given amount of warming are never fully realized (that is, C stocks fall continuously even under a small amount of warming), so are likely to markedly overestimate total soil C losses (see Methods for details). As with mechanistic models²⁶, our assumptions about the rate at which soil C responds to warming will strongly influence the magnitude of our predicted C losses (see Fig. 3b). If we make the conservative assumption that the full effects of warming are fully realized within a year, then approximately 30 ± 30 PgC would be lost from the surface soil for 1 °C of warming. Given that global average soil surface temperatures are projected to increase by around 2 °C over the next 35 years under a business-as-usual emissions scenario¹⁶, this extrapolation would suggest that warming could drive the net loss of approximately 55 ± 50 PgC from the upper soil horizon. If, as expected, this C entered the atmospheric pool, the atmospheric burden of CO₂ would increase by approximately 25 parts per million over this period.

The global extrapolation of our empirical data is broadly intended to contextualize our measured changes in soil C stocks. We stress that such statistical approaches cannot be used to project soil C losses far into the future because, unlike process-based models, they cannot capture the complex processes that govern long-term C dynamics. For example, extending the observed relationship over several centuries would lead to a global convergence of soil C stocks. Conversely, soil C stocks would increase exponentially in response to environmental cooling. Our linear extrapolation inherits weaknesses from simple single-pool models^{17,27}. However, the value of such linear approximations lies in their descriptive strengths rather than their predictive capabilities: instead of using short-term flux estimates to project long-term changes in C stocks, our approach allows the scaling of measured C differences over time frames (that is, decades) represented by the experimental studies. Our results capture the realized temperature sensitivity of current soil C stocks and can serve as a guideline (or target) for multi-pool process-based models. Specifically, these models can run forward simulations that attempt to reflect the outcomes of the warming experiments that we present. Those models that accurately capture the observed relationships between standing soil C stocks and losses under gradual step increases in global temperature are likely to be the most successful at projecting the land C–climate feedback into the future.

Our analysis reveals a number of outstanding challenges facing empiricists and modellers that limit the certainty of current land C–climate

Table 1 | List of the major remaining gaps in our understanding of the land C–climate feedback

Limitation	Consequence
Limited data on the sensitivity of C stocks in deeper soil horizons	Our global estimates of warming-induced C losses cannot represent the C losses from lower soil horizons, which are likely to enhance the magnitude of our reported effects
Uncertainty regarding current estimates of global soil C stocks	Our analysis highlights that soil C losses are highly dependent on the size of standing C stocks. Constraining our estimates of global soil C stocks should therefore be a research priority.
Limited spatial scale of field warming experiments. None of our data were collected from tropical ecosystems.	This leads to considerable uncertainty regarding the sensitivity of C stocks in many parts of the world—the tropics in particular
Practical difference between soil warming experiments and ecosystem warming experiments	Although most studies experimentally warm both the soil and plants within an ecosystem (that is, the whole ecosystem), many studies in forests are unable to capture the temperature responses of trees, which can limit comparison across ecosystem types. Depending on how trees respond to warming, soil C stock data from these sites might over-, or under-represent C losses.
Limited global data on warming-induced changes in primary productivity	Projecting these changes in soil C stocks into the future requires a mechanistic understanding of how warming affects each of the separate components of the ecosystem C cycle. The outstanding limitation, and major source of uncertainty, is the lack of data on warming-induced changes in soil C inputs at a global scale.
Our analysis does not incorporate the interactive effects of other global change drivers	Although temperature is a dominant factor driving changes in Earth system dynamics, the responses of other abiotic factors (including CO ₂ concentration, N enrichment and moisture availability) have potential to influence the strength of the relationships presented here.
Limited global data on the effects of biotic responses to warming at a global scale.	Our current understanding of global feedbacks is dominated by the physical sciences, but changes in the physiology and community compositions of organisms have been shown to have strong effects on the strength of this feedback. It is critical that we understand the importance of these effects at a global scale.
We do not know how long it takes for the effects of warming to be realized. That is, we do not know how long soil communities take to acclimatize to warming.	Extrapolating these over time necessitates that we understand the rate at which the effects of warming occur (that is, how fast soils acclimatize to warming). If the effects of warming are realized slowly, then the estimated warming-induced losses of soil C by 2050 will be towards the upper end of the uncertainty bounds that we present (approximately 200 PgC), but if soils acclimatize rapidly, then they will be towards the lower end (around 50 PgC).

Each of these limitations represents a practical challenge that can be addressed by empiricists to improve the accuracy of benchmarking estimates or to parameterize process-based models that project soil C dynamics into the future.

feedback predictions (see the list of critical research gaps in Table 1). These limitations fall into two distinct categories: more data are necessary to improve both our current global estimates of the temperature sensitivity of soil C and modelling efforts to project these soil C responses into the future. First, along with the limited spatial and temporal scale of current warming experiments, perhaps the most critical limitation to our present analysis is the paucity of information about the responses of soil C stocks at depth (below 10 cm). Although the sizes of C stocks decrease down the soil profile²⁸, any extra C losses from these deeper soil horizons will undoubtedly enhance the effects we present. Second, incorporating global soil C information into modelling frameworks requires a mechanistic understanding of how warming affects each of the individual components of the ecosystem C cycle. Now that we are beginning to generate a global picture of the temperature sensitivity of soil C losses (respiration)⁶ and total C stocks, our limited understanding of how warming influences global soil C inputs remains a major source of uncertainty for modelling efforts^{14,25}. These efforts also require more information about the interacting effects of other global change factors that may simultaneously influence soil C dynamics. This non-exclusive set of practical challenges calls for concerted, coordinated investment in multi-factor climate change experiments for an extended period of time to generate the data necessary to improve confidence in future climate projections.

In conclusion, our global compilation of experimental data allows us to see past the conflicting results from single-site studies and capture larger patterns in the sensitivity of soil C to warming. The warming-induced changes in soil C stocks reflect the net result of changes in C fluxes into and from the soil, which can augment modelling efforts to project Earth system dynamics into the future. Ultimately, our analysis provides empirical support for the long-held concern that rising temperatures stimulate the loss of soil C to the atmosphere, driving a positive land C–climate feedback that could accelerate planetary warming over the twenty-first century^{3,4}. Reductions in greenhouse gas emissions are essential if we are to avoid the most damaging effects of this feedback over the rest of the century.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 January; accepted 5 October 2016.

- Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M. & Kirk, G. J. D. Carbon losses from all soils across England and Wales 1978–2003. *Nature* **437**, 245–248 (2005).
- Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**, 165–173 (2006).
- Billings, W. D. Carbon balance of Alaskan tundra and taiga ecosystems: past, present and future. *Quat. Sci. Rev.* **6**, 165–177 (1987).
- Jenkinson, D. S., Adams, D. E. & Wild, A. Model estimates of CO₂ emissions from soil in response to global warming. *Nature* **351**, 304–306 (1991).
- Lu, M. *et al.* Responses of ecosystem carbon cycle to experimental warming: a meta-analysis. *Ecology* **94**, 726–738 (2013).
- Mahecha, M. D. *et al.* Global convergence in the temperature sensitivity of respiration at ecosystem level. *Science* **329**, 838–840 (2010).
- Ballantyne, A. P. *et al.* Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty. *Biogeosciences* **12**, 2565–2584 (2015).
- Riahi, K. *et al.* RCP 8.5: A scenario of comparatively high greenhouse gas emissions. *Climatic Change* **109**, 33–57 (2011).
- Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* **10**, 423–436 (2000).
- Crowther, T. W. & Bradford, M. A. Thermal acclimation in widespread heterotrophic soil microbes. *Ecol. Lett.* **16**, 469–477 (2013).
- Crowther, T. W. *et al.* Biotic interactions mediate soil microbial feedbacks to climate change. *Proc. Natl Acad. Sci. USA* **112**, 7033–7038 (2015).
- Arora, V. K. *et al.* Carbon–concentration and carbon–climate feedbacks in CMIP5 Earth system models. *J. Clim.* **26**, 5289–5314 (2013).
- Day, T. a., Ruhland, C. T. & Xiong, F. S. Warming increases aboveground plant biomass and C stocks in vascular-plant-dominated Antarctic tundra. *Glob. Change Biol.* **14**, 1827–1843 (2008).
- Todd-Brown, K. E. O. *et al.* Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences* **11**, 2341–2356 (2014).
- Sistla, S. A. *et al.* Long-term warming restructures Arctic tundra without changing net soil carbon storage. *Nature* **497**, 615–618 (2013).
- Bradford, M. A. *et al.* Managing uncertainty in soil carbon feedbacks to climate change. *Nat. Clim. Change* **6**, 751–758 (2016).

17. Jones, C. *et al.* Twenty-first-century compatible CO₂ emissions and airborne fraction simulated by CMIP5 Earth system models under four representative concentration pathways. *J. Clim.* **26**, 4398–4413 (2013).
18. Serreze, M. C. & Barry, R. G. Processes and impacts of Arctic amplification: a research synthesis. *Glob. Planet. Change* **77**, 85–96 (2011).
19. Koven, C. D. *et al.* A simplified, data-constrained approach to estimate the permafrost carbon–climate feedback. *Phil. Trans. R. Soc. A* **373**, 20140423 (2015).
20. Hengl, T. *et al.* SoilGrids1km—global soil information based on automated mapping. *PLoS ONE* **9**, e105992 (2014).
21. Schuur, E. A. *et al.* Climate change and the permafrost carbon feedback. *Nature* **520**, 171–179 (2015).
22. Macias-Fauria, M., Forbes, B. C., Zetterberg, P. & Kumpula, T. Eurasian Arctic greening reveals teleconnections and the potential for structurally novel ecosystems. *Nat. Clim. Change* **2**, 613–618 (2012).
23. Meehl, G. A. *et al.* Climate change projections in CESM1(CAM5) compared to CCSM4. *J. Clim.* **26**, 6287–6308 (2013).
24. Ciais, P. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) (IPCC, Cambridge Univ. Press, 2013).
25. Koven, C. D. *et al.* Controls on terrestrial carbon feedbacks by productivity versus turnover in the CMIP5 Earth System Models. *Biogeosciences* **12**, 5211–5228 (2015).
26. Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by modelling microbial processes. *Nat. Clim. Change* **3**, 909–912 (2013).
27. Georgiou, K., Koven, C. D., Riley, W. J. & Torn, M. S. Toward improved model structures for analyzing priming: potential pitfalls of using bulk turnover time. *Glob. Change Biol.* **21**, 4298–4302 (2015).
28. Conant, R. T. *et al.* Temperature and soil organic matter decomposition rates: synthesis of current knowledge and a way forward. *Glob. Change Biol.* **17**, 3392–3404 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank the Global Soil Biodiversity Initiative (GSBI) for support during this project. This project was largely funded by grants to T.W.C. from Marie Skłodowska-Curie actions, the British Ecological Society and the Yale Climate and Energy Institute. M.A.B. and W.R.W. were supported by grants from the US National Science Foundation and W.R.W. from the US Department of Energy and K.E.O.T.-B. by the Linus Pauling Distinguished Postdoctoral Fellowship programme. The experiments that produced the data were funded by grants too numerous to list here.

Author Contributions The study was conceived by T.W.C. and N.W.S., and developed by T.W.C., M.A.B., K.E.O.T.-B. and W.R.W. Statistical analysis was performed by K.E.O.T.-B., M.A.B. and B.L.S. Spatial scaling and mapping were performed by W.R.W. and C.W.R. The manuscript was written by T.W.C. with assistance from C.W.R., M.A.B., W.R.W., K.E.O.T.-B., S.D.A. and P.B.R. All other authors reviewed and provided input on the manuscript. Measurements of soil C, bulk density and geospatial data from climate change experiments around the world were provided by J.C.C., M.B.M., S.F., G.Z., A.J.B., B.E., S.R., J.H., H.L., Y.L., A.M., J.P., M.E., S.D.F., G.K.-D., C.P., P.H.T., L.L.R., E.P., S.S., J.M.L., S.D.A., K.K.T., B.E., L.N.M., I.K.S., K.S.L., Y.C., F.A.D., S.D.B., S.M., S.N., A.T.C., J.M.B., J.S.C., J.G., B.R.J., J.M., L.P.-M. and P.B.R.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.W.C. (thomas.crowther11@gmail.com).

Reviewer Information *Nature* thanks C. Jones and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data collection and standardization. Total percentage C and bulk density (BD) data ($n = 456$) were collected from each of the replicated warmed and ambient plots within 49 experimental warming studies located across North America, Europe and Asia. In several of these sites, it was not possible to access these data for deeper soil horizons. We therefore standardized collection protocols and account for the considerable variability in soil horizon depths by focusing on the top 10 cm of soil, which contains the majority of the biologically active C. Soil C stocks were then calculated for each plot (percentage C \times BD/100), and expressed as the total mass of C (kg m^{-3} soil) in the top 10 cm of each plot. Metadata for each study included the mean annual difference in soil surface temperature between warmed and ambient plots and the duration of experimental warming. These were multiplied together to generate the standardized degree-years metric (which reflects the extent and duration of warming) to permit the comparison of warming effects across sites. Other collected data included a site-specific geospatial reference (latitude and longitude), which was linked to spatially explicit estimates of soil characteristics (pH and texture using the SoilGrids database¹⁹) and climate (using the Bioclim database) following the procedure used in ref. 29. These climate and soil characteristics were then used to explore the dominant controls on the sensitivity of soil C stocks to warming across our global compilation of experimental studies.

Some of the climate change studies in this analysis contained multiple separate warming experiments. Degree-years and soil C were calculated independently for each study within a site, but all other environmental data were shared. In addition, some sites included multifactor climate change studies. For these studies, ambient and warmed plots were only compared under equivalent experimental conditions so that all other conditions remained consistent between treatments.

Statistical analysis. We fitted linear mixed models (LMMs) to evaluate the factors that correlate with the measured soil C stocks following warming. Study site was included as a random factor because clustering replicates by location could introduce spatial autocorrelation³⁰. The LMMs were fitted assuming a Gaussian error distribution in the lme4 package for the R statistical program. We constructed LMMs that included all of the putative explanatory variables to explain warmed soil C stocks including treatment variables (degrees warmed and degree-years) and environmental characteristics (standing soil C stocks (control C stocks), mean annual temperature, mean annual precipitation, pH (as H^+ ion concentration) and soil texture (with percentage clay as the representative variable)). Given the markedly different ranges in magnitudes of the explanatory variables at a global scale, variables were standardized using a z transformation before use in final models³⁰ although the response variable (soil C stock) was not standardized. Further, given a positive skew in the distributions of degrees, degree-year and control soil C, these variables were also natural-log transformed. Neither of these data transformations significantly altered the statistical outputs, so both were retained in final models. The only independent variables that were strongly correlated (pairwise coefficients > 0.4) were mean annual temperature and mean annual precipitation, and mean annual temperature and percentage clay.

Model selection was performed using maximum likelihood comparison of competing models (see Supplementary Information), using AIC and Bayesian information criterion (BIC) approaches that provided identical results. Only warming (degrees and degree-years) and standing C stock (control soil C) were retained in the most parsimonious models, (full model AIC = 381 versus final model AIC = 372; Supplementary Tables 6 and 7) and the best-fit model included an interaction between these two variables (additive model AIC = 375 versus multiplicative model AIC = 372; Supplementary Table 7). All reported P values are quasi-Bayesian, rather than the classical frequentist P values, but retain the same interpretation. We considered coefficients with $P < 0.05$ significant and coefficients with $P < 0.10$ marginally significant. Variance explained by the model was also estimated by calculating R^2 values for the minimally adequate LMM to retain the random effects structure.

The final statistical model was:

$$C_w = aC_c(\Delta T \Delta t) + bC_c + d(\Delta T \Delta t) + \varepsilon \quad (1)$$

where C_w is the C stock in the warmed treatment, C_c is the C stock in the control plots, $\Delta T \Delta t$ is the degree-years calculated by multiplying the degrees warmed by the length of the treatment, ε is the random effects term that controls for study site (see Supplementary Information) and a, b, d represent fitted coefficients for the statistical model.

Statistical model development. To scale the changes in soil C stocks, we rearranged our statistical equation to describe the relationship between standing soil C stocks (control C stocks) and warming (degree-years) over time:

$$\frac{C_w - C_c}{\Delta T \Delta t} = fC_c + g \quad (2)$$

This new model explained a considerable proportion ($R^2 = 0.606$; Supplementary Table 7) of the difference in soil C stocks between studies over treatment. This is further highlighted in Fig. 2.

We used sample-based bootstrapping (as opposed to the study-based bootstrapping in Fig. 2b) to evaluate the strength of this simple statistical relationship and to generate a margin of error for global soil C stock projections. Equation (1) was extrapolated with 95% confidence interval bounds by randomly selecting 200 samples from all studies, randomizing the control–warmed pairings and repeating the regression 1,000 times. This resulted in normally distributed parameters (see Supplementary Table 4) with the following 95% confidence interval. The intercept–slope pairs were then sampled to create the grey margin of error seen in Fig. 1.

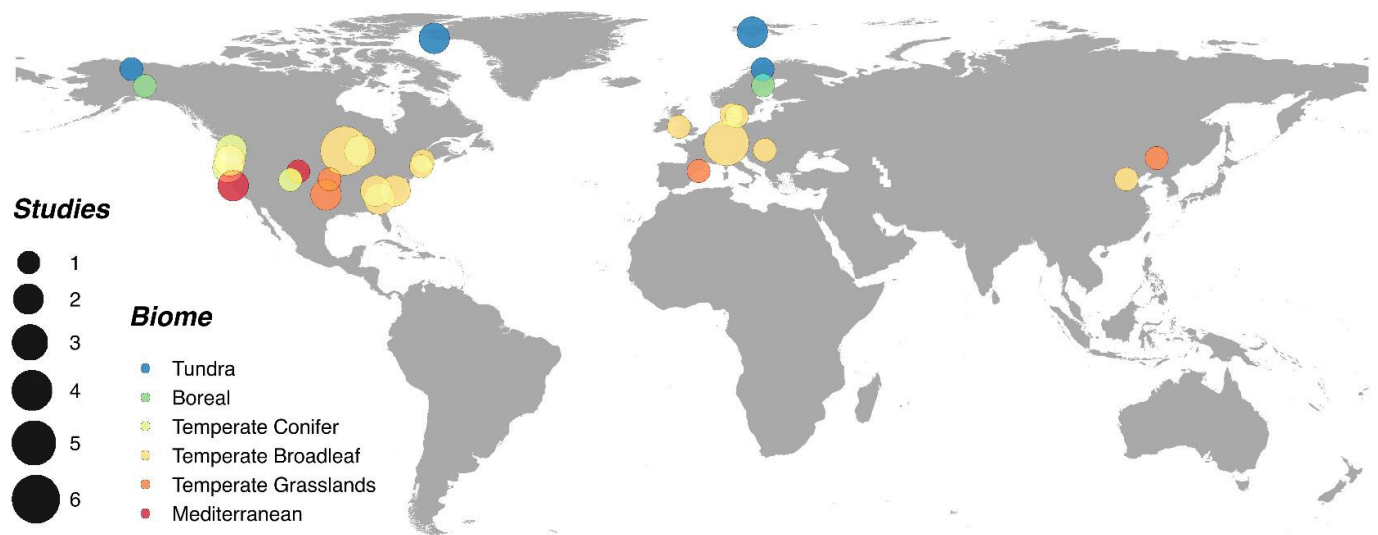
The inclusion of a linear effect of time in our analysis implicitly assumes that the effects of warming on soil C stocks are never fully realized. That is, it assumes that even a small amount of warming would continue to drive C losses indefinitely, even if the temperature were held constant. However, it is likely that C stocks would ultimately plateau (that is, acclimatize), as the full effects of warming are realized after a given period of time. As such, the assumption that the warming effects are never fully realized is likely to overestimate total soil C losses (see Fig. 3). But we do not know how long it takes for the full effects of warming to be realized (that is, how long it takes the soil to plateau/acclimatize after warming). To explore the importance of this effect-time in determining the magnitude of soil C losses in our extrapolation, we repeated the analysis across a full range of scenarios, where the effect-time of warming varied continuously. To simulate different effect-times, we successively capped the study years (or experiment duration) at 1 week, 1 month, 6 months and 1, 5, 7, 8.75, 11.6 and 17.5 years, then re-ran the linear regression described above (Equation (2) with the sample-based bootstrapping). The resulting coefficients are in Supplementary Table 4. **Extrapolation.** To estimate the changes in global soil C stocks under projected warming scenarios we applied linear changes in soil temperature that result in a mean warming of 1 °C or 2 °C by 2050 (over 35 years); this warming is spatially distributed in a manner consistent with surface soil temperature projections from a single ensemble of the Community Earth System Model (CESM) that was submitted to the CMIP5 archive under RCP8.5 run from 2005 to 2050. We estimated initial soil C stocks in the upper soil horizon (0–15 cm) from the SoilGrids 50 km² product²⁰, which was regridded using bilinear interpolation to the same spatial scale of soil surface temperature projections (roughly 1°).

The temporal extrapolations across the 35 years (until 2050) were applied separately for each of the possible effect-time scenarios described above. First, the single-time-step approach used the coefficients listed above and illustrated in Fig. 1 to generate a 95% confidence interval for projected C losses. On average, roughly 17.5 degree-years and 35 degree-years were seen cumulatively across the globe for the 1 °C and 2 °C warming scenarios, respectively. The exact warming seen by any individual grid cell was determined by the relative temperature shifts predicted by the CESM run described above. Each subsequent effect-time scenario was then extrapolated using a given time step for a forward integration where the change in soil C over that time was based on the soil C stock at the beginning and the degree-year change experienced by that site over the duration of at respective time step. For example, the 1 yr effect-time scenario used the coefficients from the analysis where experimental duration was capped at 1 yr (see Supplementary Table 4), and was extrapolated to 2050 using the sum of 35 annual time steps. The predicted soil C losses for a global average warming of 1 °C and 2 °C over 35 years, based on each of the full range of effect-time scenarios, is presented in Fig. 3b. This reveals how our assumption about effect-time influences the magnitude of our final expected C losses. Given that the effects of warming are likely to be realized within a year, we have expanded on the annual time-step option.

Code availability. The R code for the full analysis can be found in Supplementary Information.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

29. Crowther, T. W. *et al.* Mapping tree density at a global scale. *Nature* **525**, 201–205 (2015).
30. Gelman, A. Scaling regression inputs by dividing by two standard deviations. *Stat. Med.* **27**, 2865–2873 (2008).



Extended Data Figure 1 | Map of the study locations. The sizes of the points represent the number of separate warming experiments at that location and the colours indicate the biomes (as delineated by The Nature Conservancy; <http://www.nature.org>).

Unexpected diversity in socially synchronized rhythms of shorebirds

Martin Bulla¹, Mihai Valcu¹, Adriaan M. Dokter², Alexei G. Dondua³, András Kosztolányi^{4,5}, Anne L. Rutten^{1,6}, Barbara Helm⁷, Brett K. Sandercock⁸, Bruce Casler⁹, Bruno J. Ens¹⁰, Caleb S. Spiegel¹¹, Chris J. Hassell¹², Clemens Küpper¹³, Clive Minton¹⁴, Daniel Burgas^{15,16}, David B. Lank¹⁷, David C. Payer¹⁸, Egor Y. Loktionov¹⁹, Erica Nol²⁰, Eunbi Kwon²¹, Fletcher Smith²², H. River Gates^{23,24,25}, Hana Vitnerová²⁶, Hanna Prüter²⁷, James A. Johnson²⁴, James J. H. St Clair^{28,29}, Jean-François Lamarre³⁰, Jennie Rausch³¹, Jeroen Reneerkens³², Jesse R. Conklin³², Joanna Burger³³, Joe Liebezeit³⁴, Joël Bêty³⁰, Jonathan T. Coleman³⁵, Jordi Figuerola³⁶, Jos C. E. W. Hooijmeijer³², José A. Alves^{37,38}, Joseph A. M. Smith³⁹, Karel Weidinger⁴⁰, Kari Koivula⁴¹, Ken Gosbell⁴², Klaus-Michael Exo⁴³, Larry Niles⁴⁴, Laura Koloski⁴⁵, Laura McKinnon⁴⁶, Libor Praus⁴⁰, Marcel Klaassen⁴⁷, Marie-Andrée Giroux^{48,49}, Martin Sládeček⁵⁰, Megan L. Boldenow⁵¹, Michael I. Goldstein⁵², Miroslav Šálek⁵⁰, Nathan Senner^{32,53}, Nelli Rönkä⁴¹, Nicolas Lecomte⁴⁹, Olivier Gilg^{54,55}, Orsolya Vincze^{5,56}, Oscar W. Johnson⁵⁷, Paul A. Smith⁵⁸, Paul F. Woodard³¹, Pavel S. Tomkovich⁵⁹, Phil F. Battley⁶⁰, Rebecca Bentzen⁶¹, Richard B. Lanctot²⁴, Ron Porter⁶², Sarah T. Saalfeld²⁴, Scott Freeman⁶³, Stephen C. Brown²⁵, Stephen Yezerinac⁶⁴, Tamás Székely⁶⁵, Tomás Montalvo⁶⁶, Theunis Piersma^{32,67}, Vanessa Loveri⁶⁸, Veli-Matti Pakanen⁴¹, Wim Tjisen⁶⁹ & Bart Kempnaers¹

The behavioural rhythms of organisms are thought to be under strong selection, influenced by the rhythmicity of the environment^{1–4}. Such behavioural rhythms are well studied in isolated individuals under laboratory conditions^{1,5}, but free-living individuals have to temporally synchronize their activities with those of others, including potential mates, competitors, prey and predators^{6–10}. Individuals can temporally segregate their daily activities (for example, prey avoiding predators, subordinates avoiding dominants) or synchronize their activities (for example, group foraging, communal defence, pairs reproducing or caring for offspring)^{6–9,11}. The behavioural rhythms that emerge from such social synchronization and the underlying evolutionary and ecological drivers that shape them remain poorly understood^{15–7,9}. Here we investigate these rhythms in the context of biparental care, a particularly sensitive phase of social synchronization¹² where pair members potentially compromise their individual rhythms. Using data from 729 nests of 91 populations of 32 biparentally incubating shorebird species, where parents synchronize to achieve continuous coverage of developing eggs, we report remarkable within- and between-species diversity in incubation rhythms. Between species, the median length of one parent's incubation bout varied from 1–19 h, whereas period length—the time in which a parent's probability to incubate cycles once between its highest and lowest value—varied from 6–43 h. The length of incubation bouts was unrelated to variables reflecting energetic demands, but species relying on crypsis (the ability to avoid detection by other animals) had longer incubation bouts than those that are readily visible or who actively protect their nest against predators. Rhythms entrainable to the 24-h light–dark cycle were less prevalent at high latitudes and absent in 18 species. Our results indicate that even under similar environmental conditions and despite 24-h environmental cues, social synchronization can generate far more diverse behavioural rhythms than expected from studies of individuals in captivity^{5–7,9}. The risk of predation, not the risk of starvation, may be a key factor underlying the diversity in these rhythms.

Incubation by both parents prevails in almost 80% of non-passerine families¹³ and is the most common form of care in shorebirds¹⁴. Biparental shorebirds are typically monogamous¹⁵, most species lay three or four eggs in an open nest on the ground¹⁵ and cover their eggs almost continuously¹³. Pairs achieve this through synchronization of

their activities so that one of them is responsible for the nest at a given time (an incubation bout). Alternating female and male bouts generate an incubation rhythm with a specific period length (cycle of high and low probability for a parent to incubate).

We used diverse monitoring systems (Methods and Extended Data Table 1) to collect data on incubation rhythms from 91 populations of 32 shorebird species belonging to 10 genera (Fig. 1a), extracted the length of 34,225 incubation bouts from 729 nests and determined the period length for pairs in 584 nests (see Methods, Extended Data Figs 1, 2).

We found vast between- and within-species variation in incubation bout length and in period length (Figs 1–3 and Extended Data Fig. 3). Different species, but also different pairs of the same species, adopted notably different incubation rhythms, even when breeding in the same area (see, for example, incubation rhythms in Barrow, Alaska, represented by '1' in Fig. 1b, c; incubation rhythms for each nest can be found in the supplementary actograms of ref. 16). In some pairs, parents exchanged incubation duties about 20 times a day (Fig. 2a; for example *Charadrius semipalmatus*, Fig. 1b), whereas in others a single parent regularly incubated for 24 h (Fig. 2a; for example *Limnodromus scolopaceus*, Fig. 1b), with exceptional bouts of up to 50 h (supplementary actograms of ref. 16). Similarly, incubation rhythms of pairs in 22% of nests followed a strict 24-h period (Fig. 2b; for example *Tringa flavipes*, Fig. 1b), whereas the rhythms of others deviated markedly from a 24-h period (Fig. 2b) resulting in ultradian (<20 h in 12% of nests; for example *Numenius phaeopus*; Fig. 1b), free-running-like (for example *Calidris alpina*; Fig. 1b) and infradian rhythms (>28 h in 8% of nests), with some having period lengths up to 48 h (for example *Limnodromus scolopaceus*; Fig. 1b). This variation in period length partly related to the variation in bout length (Fig. 3). In the suborder Scolopaci, period length correlated positively with median bout length, but in the suborder Charadrii species with 24-h periods had various bout lengths, and species with similar bout lengths had different period lengths.

Despite substantial within-species variation, we found a strong evolutionary signal for both bout and period length with a coefficient of phylogenetic signal λ close to 1 (Extended Data Table 2). This is consistent with the notion that biological rhythms are largely genetically determined and conserved among related species^{8–10}. However, the phylogenetic effect seems unevenly distributed over the taxonomic level. Suborder explained 33% of the phenotypic variance in both bout

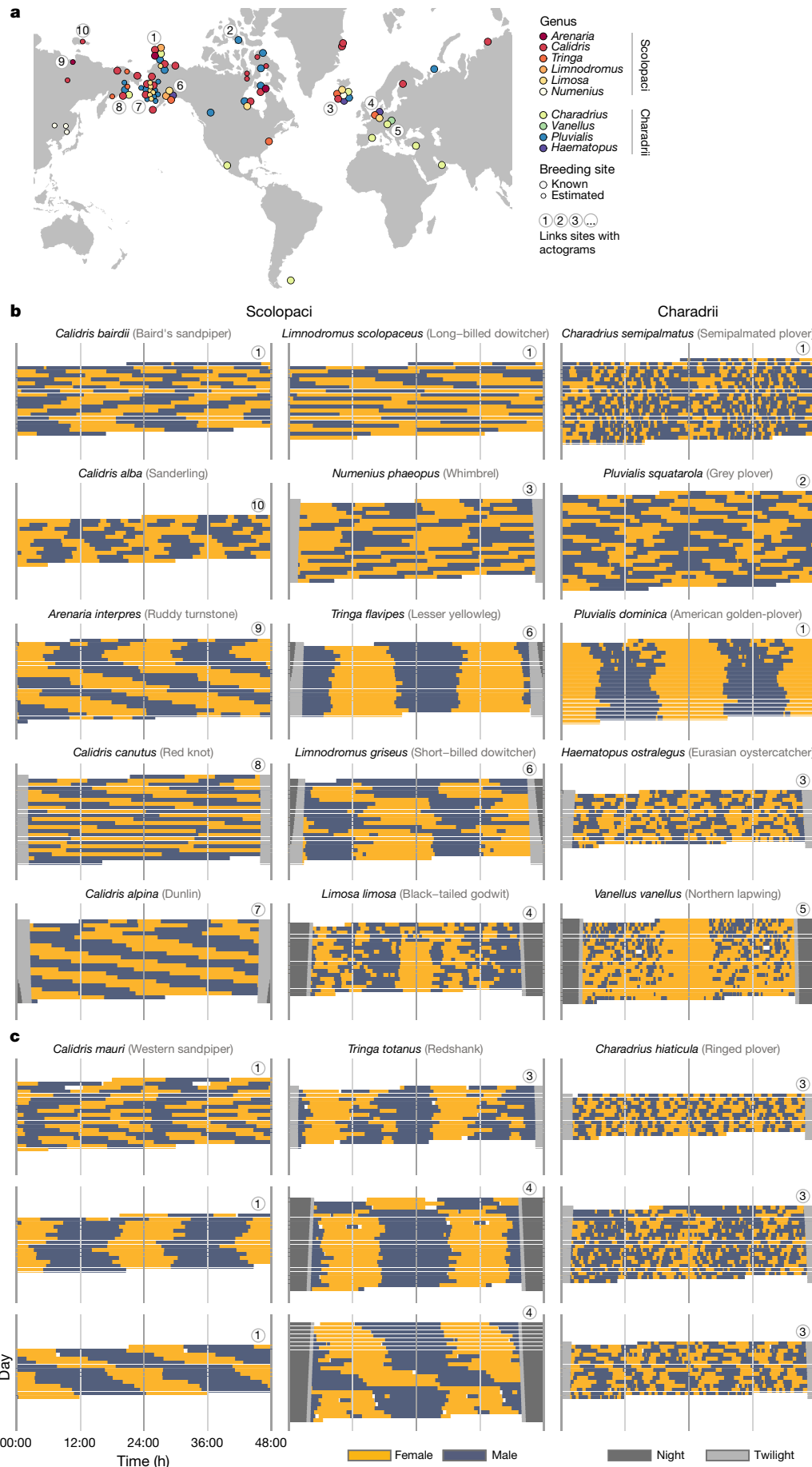


Figure 1 | Map of studied breeding sites and the diversity of shorebird incubation rhythms. **a**, Map of breeding sites with data on incubation rhythms. The colour of the dots indicates the genus (data from multiple species per genus may be available), the size of the dots refers to data quality (large dots, exact breeding site known; small dots, breeding site estimated, see Methods). For nearby or overlapping locations, the dots are scattered to increase visibility. Contours of the map were made with Natural Earth, <http://www.naturalearthdata.com>.

b, c, Illustrations of between-species diversity (**b**) and within-species diversity (**c**; note that the three rhythms for *Calidris mauri* and *Calidris hiaticula* come from the same breeding location). Each actogram depicts the bouts of female (yellow) and male (blue) incubation at a single nest over a 24-h period, plotted twice, so that each row represents two consecutive days. If present, twilight is indicated by light grey bars and corresponds to the time when the sun is between 6° and 0° below the horizon, night is indicated by dark grey bars and corresponds to the time when the sun is $>6^\circ$ below the horizon. Twilight and night are omitted in the centre of the actogram (24:00) to make the incubation rhythm visible. The circled numbers (1–10) indicate the breeding site of each pair and correspond to the circled numbers on the map in **a**.

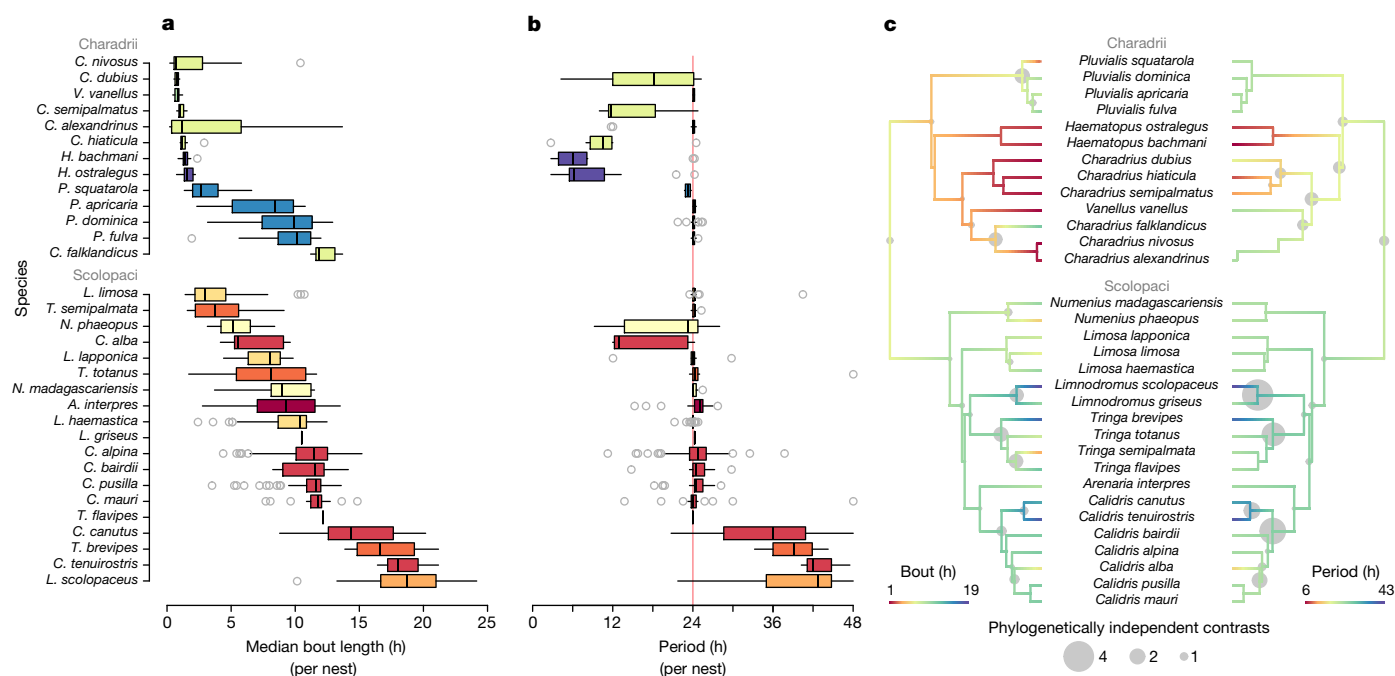


Figure 2 | Variation in incubation rhythms and its estimated evolution. **a**, **b**, Box plots are ordered by species (within suborder) from the shortest to the longest median bout length, and depict the genus (colour as in Fig. 1a), median (vertical line inside the box), 25–75th percentiles (box) 25th and 75th percentiles minus or plus $1.5 \times$ interquartile range, respectively, or the minimum and maximum value, whichever is smaller (box) outliers (circles). $n_{\text{median bout length}} = 729$ and $n_{\text{period}} = 584$

and period length, with the Scolopaci having longer incubation bouts and periods than the Charadrii (Extended Data Table 3 and Figs 2, 3). Species explained 41% of the phenotypic variation in bout length and 46% in period length, but genus explained little ($<1\%$ in both bout and period length; Extended Data Table 3), suggesting that despite a strong phylogenetic signal, these traits can rapidly diverge (Fig. 2c).

Two ecological factors may explain the observed variation in bout length. First, the ‘energetic demands hypothesis’ stipulates that the length of an incubation bout depends on the energetic state of the bird^{13,17}. This predicts that large species will have longer incubation bouts than smaller species, because they radiate less body heat per unit of mass and that incubation bouts will shorten with increasing breeding latitude, because—everything else being equal—energy

stores will deplete faster in colder environments (Extended Data Fig. 4a, b shows latitudinal cline in summer temperatures). However, bout length was unrelated to body size (Fig. 4a) and correlated positively (instead of negatively) with latitude (Fig. 4b). These correlational results across populations and species support recent experimental findings within species¹⁸ and suggest that in biparentally incubating shorebirds energetic demands are not an important ecological driver underlying variation in bout length.

An alternative explanation for variation in the length of incubation bouts relates to anti-predation strategies. Those species that rely primarily on parental crypsis (Extended Data Fig. 5a) benefit from reduced activity near the nest, because such activity can reveal the location of the nest to potential predators^{19,20}. Thus, in these species, selection will favour fewer change-overs at the nest and therefore longer incubation bouts. By contrast, species that are clearly visible when sitting on the nest or that rely on active anti-predation behaviour (Extended Data Fig. 5b), including having a partner on the watch for predators, leaving the nest long before the predator is nearby and attacking or distracting the predator¹⁵, obtain no advantage from minimizing activity. For these species, bout length can shorten, which may be advantageous for other reasons (for example, reduced need to store fat). We quantified anti-predation strategy as the distance at which the incubating parent left the nest when approached by a human (escape distance), because cryptic species stay on the nest longer (often until nearly stepped upon)¹⁵. Despite the large geographical distribution of the studied species, with related variability in the suite of predators and predation pressure²¹, and even when controlling for phylogeny (which captures much of the variation in anti-predation strategy, Extended Data Fig. 6), escape distance negatively correlated with the length of incubation bouts (Fig. 4c). This result suggests that bout length co-evolved with the anti-predation strategy.

Under natural conditions, most organisms show 24-h rhythmicity, but during the summer, when most shorebirds breed, the 24-h variation in light decreases with latitude leading to continuous polar daylight

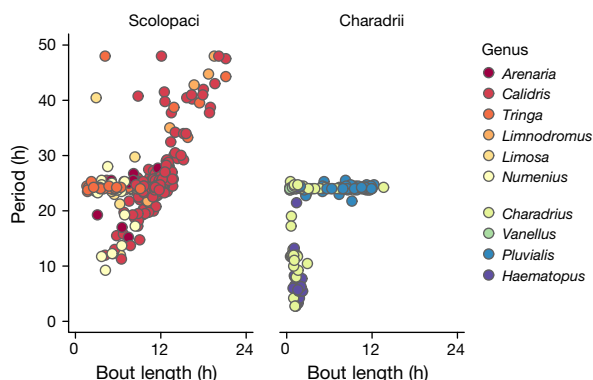


Figure 3 | Relationship between bout and period length. Each dot represents a single nest ($n = 584$ nests), colours depict the genus. In the suborder Scolopaci the median bout length and period length correlate positively ($r_{\text{Spearman}} = 0.56$, $n = 424$ nests); in the suborder Charadrii, periods longer than approximately 24 h are absent, and there is no simple relationship between bout and period length ($n = 160$ nests). For species-specific relationships see Extended Data Fig. 3.

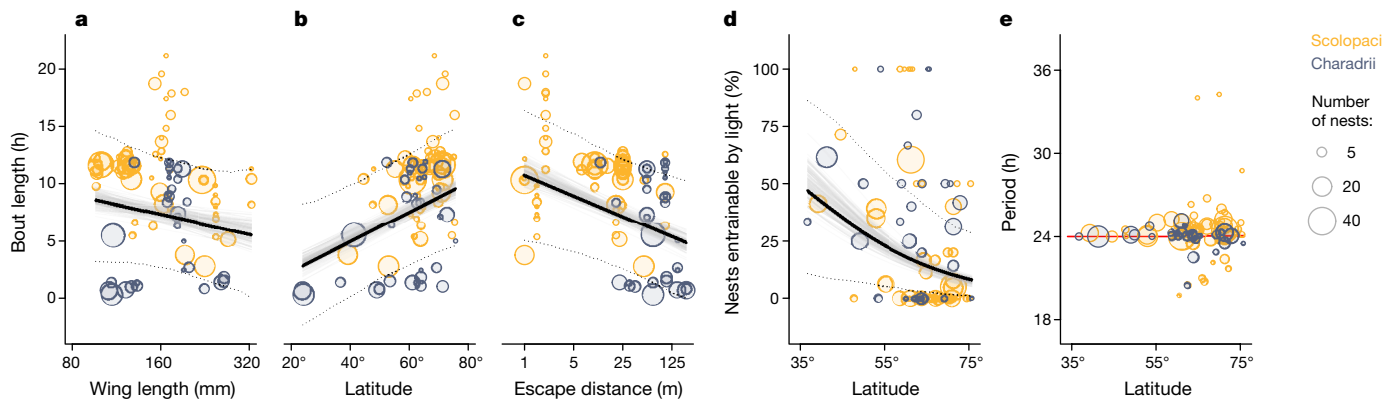


Figure 4 | Predictors of variation in incubation rhythms.

a–c, Relationships between bout length and body size, measured as female wing length (**a**), breeding latitude (**b**) and anti-predation strategy, quantified as escape distance (**c**) for $n = 729$ nests from 91 populations belonging to 32 species. **d**, The relationship between the proportion of nests with a period length that is entrainable by the 24-h light–dark cycle (that is, period lengths: 3, 6, 12, 24, or 48 h) and breeding latitude ($n = 584$ nests from 88 populations belonging to 30 species). **e**, The distribution of period length over latitude. The period was standardized to 24 h so that all 24-h harmonics are depicted as 24 h (red line) and respective deviations from each harmonic as deviations from 24 h (for example, a period of 12.5 h is depicted as 25 h). **a–e**, Each circle represents the population

in the northern-most breeding grounds²² (Extended Data Fig. 4c, d). Such reduced variation in 24-h light intensity may cause a loss of 24-h rhythmicity^{23–25}. As a consequence, circadian behavioural rhythms should exhibit a latitudinal cline²². As predicted, incubation rhythms with periods that do not follow the 24-h light–dark cycle, such as free-running-like patterns (left column in Fig. 1b), occurred more often in shorebirds breeding at higher latitudes (Fig. 4d). The absolute deviations of periods from 24 h and 24-h harmonics also increased with latitude (Fig. 4e and Extended Data Table 4). Although this supports the existence of a latitudinal cline in socially emerged behavioural rhythms²², we found a substantial number of rhythms that defy the 24-h day even at low and middle latitudes (Fig. 4d–e).

Many shorebirds predominantly use tidal habitats, at least away from their breeding ground¹⁵. To anticipate tidal foraging opportunities, these species may have activity patterns with a period length resembling the tidal period. Because changing to a different rhythm is costly²⁶, these tidal activity patterns might carry over to incubation. Although half of our species are tidal away from their breeding grounds, and some forage in tidal areas also during breeding (approximately 12% of populations), in only 5% of nests did pairs display a period length that can be entrained by the tide. Moreover, tidal species had period lengths similar to, not longer than, non-tidal ones (Extended Data Table 4). Hence, unlike the 24-h light–dark cycle, tidal life-history seems to play at best a negligible role in determining incubation rhythms.

Three main questions arise from our results. First, is variation in incubation bout length in cryptic species related to the actual predation pressure? This can be tested by comparing bout length between populations of a particular species that are exposed to different predator densities, or between years that differ in predation pressure. Second, it remains unclear how the diverse social rhythms emerge. Are these rhythms a consequence of behavioural flexibility, or a ‘fixed’ outcome of synchronization between the circadian clocks of the two individuals involved? An experimental study on ring doves (*Streptopelia risoria*) suggests that parents may even use two timers—circadian oscillation and interval timing—to determine when to incubate²⁷. Parents rapidly adjusted their schedules to phase-shifted photoperiods and their incubation rhythm ‘ran free’ in constant dim illumination (implying a circadian mechanism), whereas an experimental delay in the onset of an incubation bout did not change the length of the bout because the incubating parent refused to leave the nest until its incubation bout reached the ‘typical’ duration

(implying interval timing). Third, what are the fitness consequences for the parents of having a certain incubation rhythm? For example, the costs of having a particular incubation rhythm may be unevenly distributed between the two parents (for instance, because one parent is on incubation duty when food is more readily available, or because one parent ‘enforces’ its own rhythm at a cost to the other parent).

In conclusion, our results reveal that under natural conditions social synchronization can generate much more diverse rhythms than expected from previous work^{5–7,9,28}, and that these rhythms often defy the assumption of entrainment to the 24-h day–night cycle. Risk of predation, rather than risk of starvation, seems to have a key role in determining some of the variation in incubation rhythms. We describe this diversity in the context of biparental incubation, but diverse behavioural rhythms may also arise in many other social settings (for example, in the context of mating interactions²⁵ or vigilance behaviour during group foraging). Essentially, the reported diversity suggests that the expectation that individuals within a pair (or group) should optimize their behavioural rhythms relative to the 24-h day may be too simplistic, encouraging further study of the evolutionary ecology of plasticity in circadian clocks.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 June; accepted 21 October 2016.

Published online 23 November 2016.

- Dunlap, J. C., Loros, J. J. & DeCoursey, P. J. *Chronobiology: Biological Timekeeping* (Sinauer Associates, 2004).
- Young, M. W. & Kay, S. A. Time zones: a comparative genetics of circadian clocks. *Nat. Rev. Genet.* **2**, 702–715 (2001).
- Helm, B. & Visser, M. E. Heritable circadian period length in a wild bird population. *Proc. R. Soc. B* **277**, 3335–3342 (2010).
- Koskenvuo, M., Hublin, C., Partinen, M., Heikkilä, K. & Kaprio, J. Heritability of diurnal type: a nationwide study of 8753 adult twin pairs. *J. Sleep Res.* **16**, 156–162 (2007).
- Kronfeld-Schor, N., Bloch, G. & Schwartz, W. J. Animal clocks: when science meets nature. *Proc. R. Soc. B* **280**, 20131354 (2013).
- Bloch, G., Herzog, E. D., Levine, J. D. & Schwartz, W. J. Socially synchronized circadian oscillators. *Proc. R. Soc. B* **280**, 20130035 (2013).
- Castillo-Ruiz, A., Paul, M. J. & Schwartz, W. J. In search of a temporal niche: social interactions. *Prog. Brain Res.* **199**, 267–280 (2012).
- Mistlberger, R. E. & Skene, D. J. Social influences on mammalian circadian rhythms: animal and human studies. *Biol. Rev. Camb. Philos. Soc.* **79**, 533–556 (2004).

9. Davidson, A. J. & Menaker, M. Birds of a feather clock together—sometimes: social synchronization of circadian rhythms. *Curr. Opin. Neurobiol.* **13**, 765–769 (2003).
10. Mrosovsky, N. Locomotor activity and non-photic influences on circadian clocks. *Biol. Rev. Camb. Philos. Soc.* **71**, 343–372 (1996).
11. Regal, P. J. & Connolly, M. S. Social influences on biological rhythms. *Behaviour* **72**, 171–198 (1980).
12. Emlen, S. T. & Oring, L. W. Ecology, sexual selection, and the evolution of mating systems. *Science* **197**, 215–223 (1977).
13. Deeming, D. C. *Avian Incubation: Behaviour, Environment and Evolution* (Oxford Univ. Press, 2002).
14. Szekely, T. & Reynolds, J. D. Evolutionary transitions in parental care in shorebirds. *Proc. R. Soc. B* **262**, 57–64 (1995).
15. del Hoyo, J., Elliott, A. & Sargatal, J. *Handbook of the Birds of the World. Vol. 3. Hoatzin to Auks*. (Lynx Edicions, 1996).
16. Bulla, M. *et al.* Supporting Information for 'Unexpected diversity in socially synchronized rhythms of shorebirds'. *Open Science Framework* <http://doi.org/10.17605/OSF.IO/WXUFM> (2016).
17. Williams, J. B. in *Avian Energetics and Nutritional Ecology* (ed. C. Carey) Ch. 5, 375–416 (Chapman & Hall, 1996).
18. Bulla, M., Cresswell, W., Rutten, A. L., Valcu, M. & Kempenaers, B. Biparental incubation-scheduling: no experimental evidence for major energetic constraints. *Behav. Ecol.* **26**, 30–37 (2015).
19. Martin, T. E., Scott, J. & Menge, C. Nest predation increases with parental activity: separating nest site and parental activity effects. *Proc. R. Soc. B* **267**, 2287–2293 (2000).
20. Smith, P. A., Tulp, I., Schekkerman, H., Gilchrist, H. G. & Forbes, M. R. Shorebird incubation behaviour and its influence on the risk of nest predation. *Anim. Behav.* **84**, 835–842 (2012).
21. McKinnon, L. *et al.* Lower predation risk for migratory birds at high latitudes. *Science* **327**, 326–327 (2010).
22. Hut, R. A., Paolucci, S., Dor, R., Kyriacou, C. P. & Daan, S. Latitudinal clines: an evolutionary view on biological rhythms. *Proc. R. Soc. B* **280**, 20130433 (2013).
23. van Oort, B. E. *et al.* Circadian organization in reindeer. *Nature* **438**, 1095–1096 (2005).
24. Steiger, S. S. *et al.* When the sun never sets: diverse activity rhythms under continuous daylight in free-living arctic-breeding birds. *Proc. R. Soc. B* **280**, 20131016 (2013).
25. Lesku, J. A. *et al.* Adaptive sleep loss in polygynous pectoral sandpipers. *Science* **337**, 1654–1658 (2012).
26. Foster, R. G. & Wulff, K. The rhythm of rest and excess. *Nat. Rev. Neurosci.* **6**, 407–414 (2005).
27. Silver, R. & Bittman, E. L. Reproductive mechanisms: interaction of circadian and interval timing. *Ann. NY Acad. Sci.* **423**, 488–514 (1984).
28. Paul, M. J., Indic, P. & Schwartz, W. J. Social synchronization of circadian rhythmicity in female mice depends on the number of cohabiting animals. *Biol. Lett.* **11**, 20150204 (2015).
29. Revell, L. J. Two new graphical methods for mapping trait evolution on phylogenies. *Methods Ecol. Evol.* **4**, 754–759 (2013).
30. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).

Acknowledgements We thank all that made the data collection possible. We are grateful to W. Schwartz, E. Schlicht, W. Forstmeier, M. Baldwin, H. Fried Petersen, D. Starr-Glass and B. Bulla for comments on the manuscript and to F. Korner-Nievergelt, J. D. Hadfield, L. Z. Garamszegi, S. Nakagawa, T. Roth, N. Dochtermann, Y. Araya, E. Miller and H. Schielzeth for advice on data analysis. Data collection was supported by various institutions and people listed in supplementary data 1 at <https://osf.io/sq8gk> (ref. 16). The study was supported by the Max Planck Society (to B.K.). M.B. is a PhD student in the International Max Planck Research School for Organismal Biology.

Author Contributions M.B. and B.K. conceived the study. All authors except B.H. collected the primary data (see <https://osf.io/sq8gk>, ref. 16). M.B. coordinated the study and managed the data. M.B. and M.V. developed the methods to extract incubation. M.B. extracted bout lengths and with help from A.R. and M.V. created actograms. M.B. analysed the data with help from M.V. M.B. prepared the supporting information. M.B. and B.K. wrote the paper with input from the other authors. Except for the first, second and last author, the authors are listed alphabetically by their first name.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. (bulla.mar@gmail.com) and B.K. (b.kempenaers@orn.mpg.de).

Reviewer Information *Nature* thanks P. Bartell, C. Buck and M. Visser for their contribution to the peer review of this work.

¹Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Eberhard Gwinner Str, Seewiesen 82319, Germany. ²Computational Geo-Ecology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, Amsterdam 1098 XH, The Netherlands. ³Gatchinskaya, apartment 27, Saint Petersburg 197198, Russia. ⁴Department of Ecology, University of Veterinary Medicine Budapest, Rottenbiller u. 50, Budapest H-1077, Hungary. ⁵MTA-DE 'Lendület' Behavioural Ecology

Research Group, Department of Evolutionary Zoology, University of Debrecen, Egyetem tér 1, Debrecen H-4032, Hungary. ⁶Apiloa GmbH, Starnberg 82319, Germany. ⁷Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Graham Kerr Building, Glasgow G12 8QQ, UK. ⁸Division of Biology, Kansas State University, 116 Ackert Hall, Manhattan, Kansas 66506-4901, USA. ⁹PO Box 1094, Fallon, Nevada 89407, USA. ¹⁰Coastal Ecology Team, Sovon Dutch Centre for Field Ornithology, PO Box 59, Den Burg 1790 AB, Texel, The Netherlands. ¹¹Division of Migratory Birds, Northeast Region, US Fish and Wildlife Service, 300 Westgate Center Dr, Hadley, Massachusetts 01035, USA. ¹²Global Flyway Network, PO Box 3089, Broome, Western Australia 6725, Australia. ¹³Institute of Zoology, University of Graz, Universitätsplatz 2, 8010 Graz, Austria. ¹⁴Victorian Wader Study group, 165 Dalgetty Road, Beaumaris, Melbourne, Victoria 3193, Australia. ¹⁵Department of Forest Sciences, University of Helsinki, PO Box 27, Helsinki FI-00014, Finland. ¹⁶Department of Biological and Environmental Sciences, University of Jyväskylä, PO Box 35, Jyväskylä FI-40014, Finland. ¹⁷Department of Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada. ¹⁸Alaska Region, US National Park Service, 240 W 5th Ave, Anchorage, Alaska 99501, USA. ¹⁹State Lab for Photon Energetics, Bauman Moscow State Technical University, 2nd Baumanskaya St, 5-1, Moscow 105005, Russia. ²⁰Biology Department, Trent University, 2140 East Bank Drive, Peterborough, Ontario K9L 0G2, Canada. ²¹Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute and State University, 310 West Campus Drive, Blacksburg, Virginia 24061, USA. ²²Center for Conservation Biology, College of William & Mary and Virginia Commonwealth University, PO Box 8795, Williamsburg, Virginia 23187, USA. ²³Pacific Ecological Services, 17520 Snow Crest Lane, Anchorage, Alaska, 99516, USA. ²⁴Migratory Bird Management, US Fish and Wildlife Service, 1011 East Tudor Road, Anchorage, Alaska 99503, USA. ²⁵Shorebird Recovery Program, Manomet, PO Box 545, Saxtons River, Vermont 05154, USA. ²⁶Faculty of Science, Charles University in Prague, Albertov 6, Praha 128 43, Czech Republic. ²⁷Department of Wildlife Diseases, Leibniz Institute for Zoo- and Wildlife Research, Alfred-Kowalke-Straße 17, Berlin 10315, Germany. ²⁸Biodiversity Lab, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA1 7AY, UK. ²⁹Centre for Evolutionary Biology, School of Biology, University of Western Australia, Stirling Highway, Crawley, Western Australia 6009, Australia. ³⁰Département de biologie, chimie et géographie et Centre d'études nordiques (CEN), Université du Québec à Rimouski, 300 allée des Ursulines, Rimouski, Quebec G5L 3A1, Canada. ³¹Canadian Wildlife Service, Environment and Climate Change Canada, PO Box 2310, 5019—52nd Street, 4th Floor, Yellowknife, Northwest Territories X1A 2P7, Canada. ³²Conservation Ecology Group, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Nijenborgh 7, Groningen 9747 AG, The Netherlands. ³³Division of Life Sciences, Rutgers University, 604-Allison Road, Piscataway, New Jersey 08854-8082, USA. ³⁴Audubon Society of Portland, 5151 NW Cornell Road, Portland, Oregon 97210, USA. ³⁵Queensland Wader Study Group, 22 Parker Street, Brisbane, Queensland 4128, Australia. ³⁶Department of Wetland Ecology, Doñana Biological Station (CSIC), Av. Américo Vespucio, s/n, Seville 41092, Spain. ³⁷Centre for Environmental and Marine Studies (CESAM), Department of Biology, University of Aveiro, Campus de Santiago, Aveiro 3810-193, Portugal. ³⁸South Iceland Research Centre, University of Iceland, Fjolheimar, Selfoss 800, Iceland. ³⁹LJ Niles Associates, PO Box 784, Cape May, New Jersey 08204, USA. ⁴⁰Department of Zoology and Laboratory of Ornithology, Palacký University, 17. Listopadu 50, Olomouc 771 46, Czech Republic. ⁴¹Department of Ecology, University of Oulu, PO Box 3000, Oulu 90014, Finland. ⁴²Australasian Wader Studies Group, 1/19 Baldwin Road, Blackburn, Melbourne, Victoria 3130, Australia. ⁴³Institute of Avian Research, Vogelwarte Helgoland, An der Vogelwarte 21, Wilhelmshaven D-26386, Germany. ⁴⁴LJ Niles Associates, 109 Market Lane, Greenwich, Connecticut 08323, USA. ⁴⁵Environmental and Life Sciences, Trent University, 1600 West Bank Dr, Peterborough, Ontario K0L 0G2, Canada. ⁴⁶Bilingual Biology Program, York University Glendon Campus, 2275 Bayview Avenue, Toronto, Ontario M4N 3M6, Canada. ⁴⁷Centre for Integrative Ecology, Deakin University, 75 Pigdons Road, Warrn Ponds, Geelong, Victoria 3216, Australia. ⁴⁸Canada Research in Northern Biodiversity and Centre d'Études Nordiques, Université du Québec à Rimouski, 300, Allée des Ursulines, Rimouski, Quebec G5L 3A8, Canada. ⁴⁹Canada Research in Polar and Boreal Ecology and Centre d'Études Nordiques, Université de Moncton, 18 avenue Antonine-Maillet, Moncton, New Brunswick E4K 1A6, Canada. ⁵⁰Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamýcká 1176, Suchbát, Prague 16521, Czech Republic. ⁵¹Department of Biology and Wildlife, University of Alaska Fairbanks, PO Box 756100, Fairbanks, Alaska 99775-6100, USA. ⁵²Alaska Coastal Rainforest Center, University of Alaska Southeast, 11120 Glacier Hwy, Juneau, Alaska 99801, USA. ⁵³Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca 14850, USA. ⁵⁴Equipe Ecologie Evolution, UMR 6282 Biogéosciences, Université de Bourgogne Franche Comté, 6 Bd Gabriel, Dijon 21000, France. ⁵⁵Groupe de Recherche en Ecologie Arctique, 16 Rue de Vernot, Francheville 21440, France. ⁵⁶Evolutionary Ecology Group, Hungarian Department of Biology and Ecology, Babeş-Bolyai University, Clinicilor 5-7, Cluj-Napoca RO-400006, Romania. ⁵⁷Department of Ecology, Montana State University, Bozeman, Montana 59717, USA. ⁵⁸Wildlife Research Division, Environment and Climate Change Canada, 1125 Colonel By Dr, Ottawa, Ontario K1A 0H3, Canada. ⁵⁹Zoological Museum, Lomonosov Moscow State University, Bolshaya Nikitskaya St, 6, Moscow 125009, Russia. ⁶⁰Institute of Agriculture & Environment, Massey University, Private Bag 11 222, Palmerston North 4442, New Zealand. ⁶¹Arctic Beringia Program, Wildlife Conservation Society, 925 Schloesser Dr, Fairbanks, Alaska 99709, USA. ⁶²Delaware Bay Shorebird Project, Ambler, Pennsylvania 19002, USA. ⁶³Arctic National Wildlife Refuge, US Fish and Wildlife Service, 101 12th Ave, Fairbanks, Alaska 99701, USA. ⁶⁴Fieldday Consulting, Surrey, British Columbia V4N 6M5, Canada. ⁶⁵Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK. ⁶⁶Servei de Vigilància i Control de Plagues Urbanes, Agència de Salut Pública de Barcelona, Av. Príncep d'Astúries 63, Barcelona 8012, Spain. ⁶⁷NIOZ Royal Netherlands Institute for Sea Research, Department of Coastal Systems and Utrecht University, PO Box 59, Den Burg 1790 AB, Texel, The Netherlands. ⁶⁸Migratory Bird and Habitat Program, US Fish and Wildlife Service, 911 NE 11th Avenue, Portland, Oregon 97232, USA. ⁶⁹Poelweg 12, Westerland 1778 KB, The Netherlands.

METHODS

Recording incubation. Incubation data were obtained between 1994 and 2015, for as many shorebird species ($n = 32$) and populations ($n = 91$) as possible (that is, no statistical methods were used to predetermine sample size), using six methods (for specifications of the equipment see Extended Data Table 1). (1) In 261 nests, a radio frequency identification reader (RFID) registered the presence of tagged parents at the nest. The passive integrated tag was either embedded in a plastic flag^{31,32}, with which the parents were banded, or glued to the tail feathers³³. In 200 nests the RFID was combined with a temperature probe placed between the eggs. The temperature recordings allowed us to identify whether a bird was incubating even in the absence of RFID readings; an abrupt change in temperature marked the start or end of incubation³¹. (2) For 396 nests, light-loggers were mounted to the plastic flag or a band that was attached to the bird's leg^{34,35}. The logger recorded maximum light intensity (absolute or relative) for a fixed sampling interval (2–10 min). An abrupt change in light intensity (as opposed to a gradual change caused, for example, by twilight) followed by a period of low or high light intensity marked the start or end of the incubation period (Extended Data Fig. 2). (3) For nine nests a GPS tag, mounted on the back of the bird, recorded the position of the bird³⁶. The precision of the position depends on cloud cover and sampling interval³⁶. Hence, to account for the imprecision in GPS positions, we assumed incubation whenever the bird was within 25 m of the nest (Extended Data Fig. 2b). (4) At three nests automated receivers recorded signal strength of a radio tag attached to the rump of a bird; whenever a bird incubated, the strength of the signal remained constant²⁴ (supplementary actograms; pages 257–259 of ref. 16). (5) At 53 nests video cameras were used to identify the incubating parents. (6) Eight nests were continuously observed. In (5) and (6) parent identification was based on plumage, colour rings or radio tag. In one of the populations, three different methods were used, in seven populations representing seven species two methods were used. In one nest, two methods were used simultaneously (Extended Data Fig. 2b).

Extraction of incubation bouts. An incubation bout was defined as the total time allocated to a single parent (that is, the time between the arrival of a parent at and its departure from the nest followed by incubation of its partner). Bout lengths were only extracted if at least 24 h of continuous recording was available for a nest; in such cases, all bout lengths were extracted. For each nest, we transformed the incubation records to local time as UTC time + (longitude of the nest/15). Incubation bouts from RFIDs, videos and continuous observations were mostly extracted by an R script and the results verified by visualizing the extracted and the raw data^{16,31,37,38}; otherwise, M.B. extracted the bouts manually from plots of raw data^{39,40} (plots of raw data and extracted bouts for all nests are in the supplementary actograms of ref. 16; the actograms were generated by the ggplot and xyplot functions from the ggplot2 and lattice R-packages^{41–43}). Whenever the start or end of a bout was unclear, we classified these bouts as uncertain (see next paragraph for treatment of uncertain bouts). In case of light-logger data, the light recordings before and after the breeding period, when the birds were definitely not incubating, helped to distinguish incubation from non-incubation. Whenever an individual tagged with a light-logger nested in an environment where the sun was more than 6° below the horizon for part of a day (that is, night), we assumed an incubation bout when the individual started incubating before the night started and ended incubating after the night ended. When different individuals incubated at the beginning versus at the end of the night, we either did not quantify these bouts or we indicated the possible time of exchange (based on trends in previous exchanges), but classified these bouts as uncertain (see supplementary actograms¹⁶). In total, we extracted 34,225 incubation bouts.

The proportion of uncertain bouts within nests had a distribution skewed towards zero (median = 0%, range, 0–100%, $n = 729$ nests), and so did the median proportion of uncertain bouts within populations (median = 2%, range, 0–74%, $n = 91$ populations). Excluding the uncertain bouts did not change our estimates of median bout length (Pearson's correlation coefficient for median bout length based on all bouts and without uncertain bouts: $r = 0.96$, $n = 335$ nests with both certain and uncertain bouts). Hence, in further analyses all bouts were used to estimate median bout length.

Note that in some species sexes consistently differed in bout length (Fig. 1b, for example, *Vanellus vanellus*). As these differences are small compared to the between-species differences and because in 27 nests (of 8 species) the sex of the parents was unknown, we used median bout length independent of sex in this study.

Extraction of period length. The method used for extracting the period length of incubation rhythm for each nest is described in the Extended Data Fig. 1.

Extraction of entrainable periods. We classified 24-h periods and periods with 24-h harmonics (that is, 3, 6, 12, 48 h) as strictly entrainable by 24-h light fluctuations ($n = 142$ nests out of 584). Including nearest adjacent periods (± 0.25 h) increased the number of nests with entrainable periods ($n = 277$), but results of statistical analyses remained quantitatively similar. We consider periods and

harmonics of 12.42 h (that is, 3.1, 6.21, 12.42, 24.84 h) as strictly entrainable by tide. However, because the periods in our data were extracted in 0.25-h intervals (Extended Data Fig. 1), we classified periods of 3, 6.25, 12.5, 24.75 h (that is, those closest to the strict tide harmonics) as entrainable by tide ($n = 32$ nests out of 584). Including also the second nearest periods (that is, 3.25, 6, 12.25, 25) increased the number of nests entrainable by tide to $n = 55$.

Population or species life-history traits. For 643 nests, the exact breeding location was known (nests or individuals were monitored at the breeding area). For the remaining 86 nests (from 27 populations representing 8 species, where individuals were tagged with light-loggers at the wintering area), the breeding location was roughly estimated from the recorded 24-h variation in daylight, estimated migration tracks, and the known breeding range of the species^{44–51}. One exact breeding location was in the Southern Hemisphere, so we used absolute latitude in analyses. Analyses without populations with estimated breeding-location or without the Southern Hemisphere population generated quantitatively similar estimates as the analyses on full data.

For each population, body size was defined as mean female wing length⁵², either for individuals measured at the breeding area or at the wintering area. In case no individuals were measured, we used the mean value from the literature (see open access data for specific values and references⁵³).

Anti-predation strategy was assessed by estimating the escape distance of the incubating bird when a human approached the nest, because species that are cryptic typically stay on the nest much longer than non-cryptic species, sometimes until nearly stepped upon^{48,54}. Escape distance was obtained for all species. Forty-four authors of this paper estimated the distance (in metres) for one or more species based on their own data or experience. For ten species, we also obtained estimates from the literature⁴⁸. We then used the median 'estimated escape distance' for each species. In addition, for 13 species we obtained 'true escape distance'. Here, the researcher approached a nest (of known position) and either estimated his distance to the nest or marked his position with GPS when the incubating individual left the nest. For each GPS position, we calculated the Euclidian distance from the nest. In this way we obtained multiple observations per nest and species, and we used the median value per species (weighted by the number of estimates per nest) as the true escape distance. The species' median estimated escape distance was a good predictor of the true escape distance (Pearson's correlation coefficient: $r = 0.89$, $n = 13$ species). For analysis, we defined the escape distance of a species as the median of all available estimates.

For each species, we determined whether it predominantly uses a tidal environment outside its breeding ground, that is, has tidal versus non-tidal life history (based on refs 48, 50, 51). For each population with exact breeding location, we scored whether tidal foraging habitats were used by breeding birds for foraging (for three populations this information was unknown)⁵³. For all populations with estimated breeding location we assumed, based on the estimated location and known behaviour at the breeding grounds, no use of tidal habitat.

Statistical analyses. Unless specified otherwise, all analyses were performed on the nest level using median bout length and extracted period length.

We used phylogenetically informed comparative analyses to assess how evolutionary history constrains the incubation rhythms (estimated by Pagel's λ coefficient of phylogenetic signal^{55,56}) and to control for potential non-independence among species due to common ancestry. This method explicitly models how the covariance between species declines as they become more distantly related^{55,57,58}. We used the Hackett⁵⁹ backbone phylogenetic trees available at <http://birdtree.org> (ref. 60), which included all but one species (*Charadrius nivosus*) from our dataset. Following a subsequent taxonomic split⁶¹, we added *C. nivosus* to these trees as a sister taxon of *C. alexandrinus*. Phylogenetic uncertainty was accounted for by fitting each model with 100 phylogenetic trees randomly sampled from 10,000 phylogenies at <http://birdtree.org> (ref. 60).

The analyses were performed with Bayesian phylogenetic mixed-effect models (Fig. 4 and Extended Data Tables 2, 4) and the models were run with the MCMCglmm function from the R package MCMCglmm⁶². In all models, we also accounted for multiple sampling within species and breeding site (included as random effects). In models with a Gaussian response variable, an inverse-gamma prior with shape and scale equal to 0.001 was used for the residual variance (that is, variance set to 1 and the degree of belief parameter to 0.002). In models with binary response variables, the residual variance was fixed to 1. For all other variance components the parameter-expanded priors were used to give scaled F -distributions with numerator and denominator degrees of freedom set to 1 and a scale parameter of 1,000. Model outcomes were insensitive to prior parameterization. The MCMC chains ran for 2,753,000 iterations with a burn-in of 3,000 and a thinning interval of 2,500. Each model generated approximately 1,100 independent samples of model parameters (Extended Data Tables 2, 4). Independence of samples in the Markov chain was assessed by tests for autocorrelation between samples and by using graphic diagnostics.

First, we used MCMCglmm to estimate Pagel's λ (phylogenetic signal) for bout and period length (Gaussian), and to show that our estimates of these two incubation variables were independent of how often the incubation behaviour was sampled ('sampling' in min, ln-transformed; Extended Data Table 2). Hence, in subsequent models, sampling was not included.

Then, we used MCMCglmm to model variation in bout length and period length (Extended Data Table 4). Bout length was modelled as a continuous response variable and latitude (in degrees, absolute), female wing length (mm, ln-transformed) and approach distance (m, ln-transformed) as continuous predictors. Predictors had low collinearity (at nest, population and species level; all Pearson or Spearman correlation coefficients $|r| < 0.28$). To test for potential entrainment to 24-h, period length was modelled as a binary response variable (1 = rhythms with period of 3, 6, 12, 24, or 48 h; 0 = rhythms with other periods) and latitude as a continuous predictor. To test how circadian period varies with latitude or life history, the period was transformed to deviations from 24 h and 24-h harmonics and scaled by the time span between the closest harmonic and the closest midpoint between two harmonics. For example, a 42-h period deviates by -6 h from 48 h (the closest 24-h harmonic) and hence -6 h was divided by 12 h (the time between 36 h—the midpoint of two harmonics—and 48 h—the closest harmonic). This way the deviations spanned from -1 to 1 with 0 representing 24 h and its harmonics. The absolute deviations were then modelled as a continuous response variable and latitude as continuous predictor. The deviations were also modelled as a continuous response and species life history (tidal or not) as categorical predictor.

In all models the continuous predictor variables were centred and standardized to a mean of zero and a standard deviation of one.

We report model estimates for fixed and random effects, as well as for Pagel's λ , by the modes and the uncertainty of the estimates by the highest posterior density intervals (referred to as 95% CI) from the joint posterior distributions of all samples from the 100 separate runs, each with 1 of the 100 separate phylogenetic trees from <http://birdtree.org> (ref. 60).

To help interpret the investigated relationships we assessed whether incubation rhythms evolved within diverged groups of species by plotting the evolutionary tree of the incubation rhythm variables (Fig. 2c), as well as of the predictors (Extended Data Fig. 6).

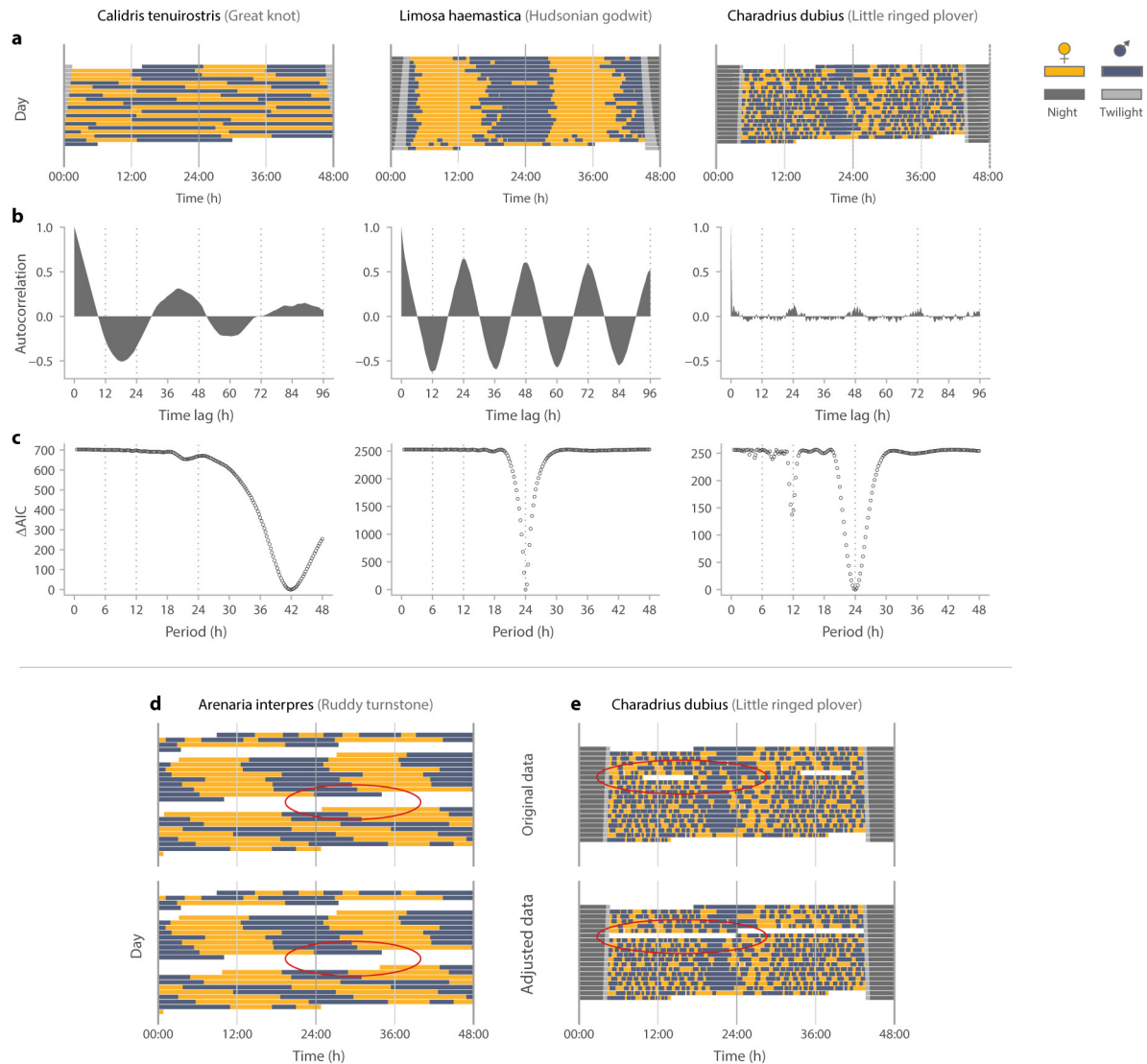
The source of phylogenetic constraint in bout and period length was investigated by estimating the proportion of phenotypic variance explained by suborder, genus and species (Extended Data Table 3). The respective mixed models were also specified with MCMCglmm⁶² using the same specifications as in the phylogenetic models. Because suborder contained only two levels, we first fitted an intercept mixed model with genus, species, and breeding site as random factors, and used it to estimate the overall phenotypic variance. We then entered suborder as a fixed factor and estimated the variance explained by suborder as the difference between the total variance from the first and the second model. To evaluate the proportion of the variance explained by species, genus and breeding site, we used the estimates from the model that included suborder.

R version 3.1.1 (ref. 63) was used for all statistical analyses.

Code availability. All statistical analyses, figures, and the supplementary actograms are replicable with the open access information, including computer software and code for R, available at the Open Science Framework, <https://osf.io/wxufm/> (ref. 16).

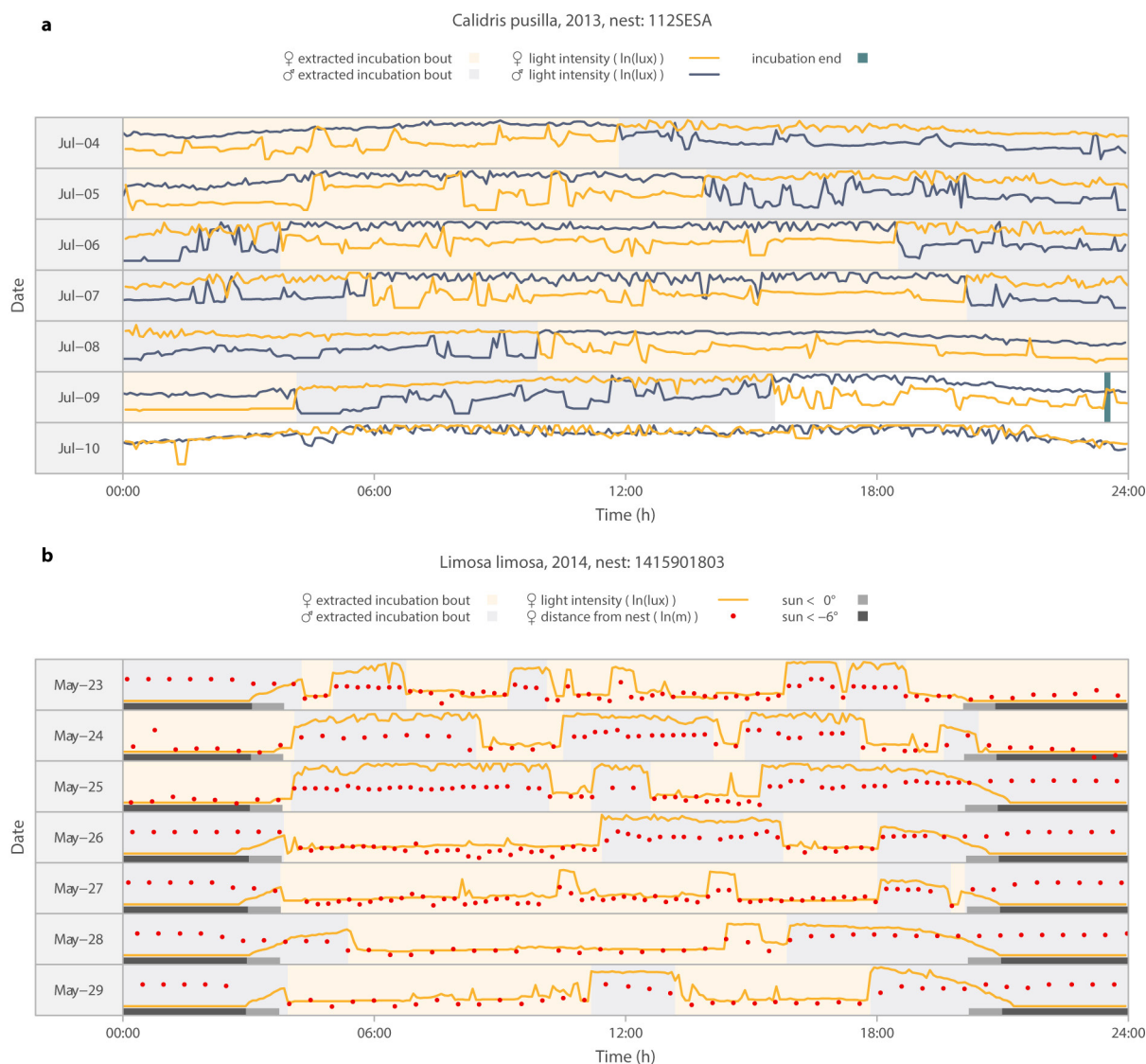
Data availability. Primary and extracted data that support the findings of this study are freely available from the Open Science Framework, <https://osf.io/wxufm/> (ref. 16). Source data for Figs 1–4 are provided with the paper.

31. Bulla, M., Valcu, M., Rutten, A. L. & Kempenaers, B. Biparental incubation patterns in a high-Arctic breeding shorebird: how do pairs divide their duties? *Behav. Ecol.* **25**, 152–164 (2014).
32. Reneerkens, J., Grond, K., Schekkerman, H., Tulp, I. & Piersma, T. Do uniparental sanderlings *Calidris alba* increase egg heat input to compensate for low nest attentiveness? *PLoS One* **6**, e16834 (2011).
33. Kosztolányi, A. & Székely, T. Using a transponder system to monitor incubation routines of Snowy Plovers. *J. Field Ornithol.* **73**, 199–205 (2002).
34. Conklin, J. R. & Battley, P. F. Attachment of geolocators to bar-tailed godwits: a tibia-mounted method with no survival effects or loss of units. *Wader Study Group Bull.* **117**, 56–58 (2010).
35. Burger, J., Niles, L. J., Porter, R. R. & Dey, A. D. Using geolocator data to reveal incubation periods and breeding biology in Red Knots *Calidris canutus rufa*. *Wader Study Group Bull.* **119**, 26–36 (2012).
36. Bouten, W., Baaij, E. W., Shamoun-Baranes, J. & Camphuysen, K. C. J. A flexible GPS tracking system for studying bird behaviour at multiple scales. *J. Ornithol.* **154**, 571–580 (2012).
37. Bulla, M. R-SCRIPT and EXAMPLE DATA to extract incubation from temperature measurements. Version 1. *figshare* <https://dx.doi.org/10.6084/m9.figshare.1037545.v1> (2014).
38. Bulla, M. R-SCRIPT and EXAMPLE DATA to extract incubation bouts from continuous RFID and video data. Version 1. *figshare* <https://dx.doi.org/10.6084/m9.figshare.1533278.v1> (2015).
39. Bulla, M. Example of how to manually extract incubation bouts from interactive plots of raw data—R-CODE and DATA. Version 1. *figshare* <https://dx.doi.org/10.6084/m9.figshare.2066784.v1> (2016).
40. Bulla, M. Procedure for manual extraction of incubation bouts from plots of raw data.pdf. Version 1. *figshare* <https://dx.doi.org/10.6084/m9.figshare.2066709.v1> (2016).
41. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer 2009).
42. Sarkar, D. & Andrews, F. *latticeExtra: Extra Graphical Utilities Based on Lattice*. R package version 0.6-24 <http://CRAN.R-project.org/package=latticeExtra> (2012).
43. Sarkar, D. *Lattice: Multivariate Data Visualization with R* (Springer, 2008).
44. Lisovski, S. Geolocator-ArcticWader-BreedingSiteEstimation. Version 2015-08-05. *GitHub repository* <https://github.com/slisovski/Geolocator-ArcticWader-BreedingSiteEstimation> (2015).
45. Lisovski, S. & Hahn, S. GeoLight—processing and analysing light-based geolocator data in R. *Methods Ecol. Evol.* **3**, 1055–1059 (2012).
46. Lisovski, S. et al. Geolocation by light: accuracy and precision affected by environmental factors. *Methods Ecol. Evol.* **3**, 603–612 (2012).
47. Conklin, J. R., Battley, P. F., Potter, M. A. & Fox, J. W. Breeding latitude drives individual schedules in a trans-hemispheric migrant bird. *Nat. Commun.* **1**, 67 (2010).
48. Poole, A. *The Birds of North America* (Cornell Laboratory of Ornithology, 2005).
49. Lappo, E., Tomkovich, P. & Syroechkovskiy, E. *Atlas of Breeding Waders in the Russian Arctic* (UF Ofsetnaya Pechka, 2012).
50. Chandler, R. J. *Shorebirds of the Northern Hemisphere* (Christopher Helm, 2009).
51. Brazil, M. *Birds of East Asia: Eastern China, Taiwan, Korea, Japan, and Eastern Russia* (Christopher Helm, 2009).
52. Dale, J. et al. Sexual selection explains Rensch's rule of allometry for sexual size dimorphism. *Proc. R. Soc. B* **274**, 2971–2979 (2007).
53. Bulla, M. et al. Supplementary Data 3—Study sites: location, population wing length, monitoring method, tide. Version 11. *figshare* <https://dx.doi.org/10.6084/m9.figshare.1536260.v11> (2016).
54. Cramp, S. *Handbook of the Birds of Europe, the Middle East, and North Africa: The Birds of the Western Palearctic Volume III: Waders to Gulls* (Oxford Univ. Press, 1990).
55. Freckleton, R. P., Harvey, P. H. & Pagel, M. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* **160**, 712–726 (2002).
56. Pagel, M. Inferring evolutionary processes from phylogenies. *Zool. Scr.* **26**, 331–348 (1997).
57. Martins, E. P. & Hansen, T. F. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* **149**, 646–667 (1997).
58. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).
59. Hackett, S. J. et al. A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763–1768 (2008).
60. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
61. Küpper, C. et al. Kentish versus snowy plover: phenotypic and genetic analyses of *Charadrius alexandrinus* reveal divergence of Eurasian and American subspecies. *Auk* **126**, 839–852 (2009).
62. Hadfield, J. D. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J. Stat. Softw.* **33**, 1–22 (2010).
63. R-Core-Team. R: A Language and Environment for Statistical Computing. Version 3.1.1. *R Foundation for Statistical Computing* <http://www.R-project.org/> (2014).
64. Anderson, D. R. *Model Based Inference in the Life Sciences: A Primer on Evidence* (Springer, 2008).
65. Hijmans, R. J. raster: Geographic data analysis and modeling. R package version 2.3-24. <http://CRAN.R-project.org/package=raster> (2015).
66. Bivand, R. & Lewin-Koh, N. maptools: Tools for reading and handling spatial objects. R package version 0.8-30. <http://CRAN.R-project.org/package=maptools> (2014).
67. Revell, L. J. in *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology* (ed. L. Z. Garamszegi) Ch. 4, 77–103 (Springer, 2014).
68. Johnson, O. W. et al. Tracking Pacific golden-plovers *Pluvialis fulva*: transoceanic migrations between non-breeding grounds in Kwajalein, Japan and Hawaii and breeding grounds in Alaska and Chukotka. *Wader Study* **122**, 13–20 (2015).
69. Kosztolányi, A., Cuthill, I. C. & Székely, T. Negotiation between parents over care: reversible compensation during incubation. *Behav. Ecol.* **20**, 446–452 (2009).
70. St Clair, J. J. H., Herrmann, P., Woods, R. W. & Székely, T. Female-biased incubation and strong diel sex-roles in the two-banded plover *Charadrius falklandicus*. *J. Ornithol.* **151**, 811–816 (2010).
71. Spiegel, C. S., Haig, S. M., Goldstein, M. I. & Huso, M. Factors affecting incubation patterns and sex roles of black oystercatchers in Alaska. *Condor* **114**, 123–134 (2012).
72. Praus, L. & Weidinger, K. Predators and nest success of sky larks *Alauda arvensis* in large arable fields in the Czech Republic. *Bird Study* **57**, 525–530 (2010).
73. Orme, D. et al. caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5.2. <http://CRAN.R-project.org/package=caper> (2013).



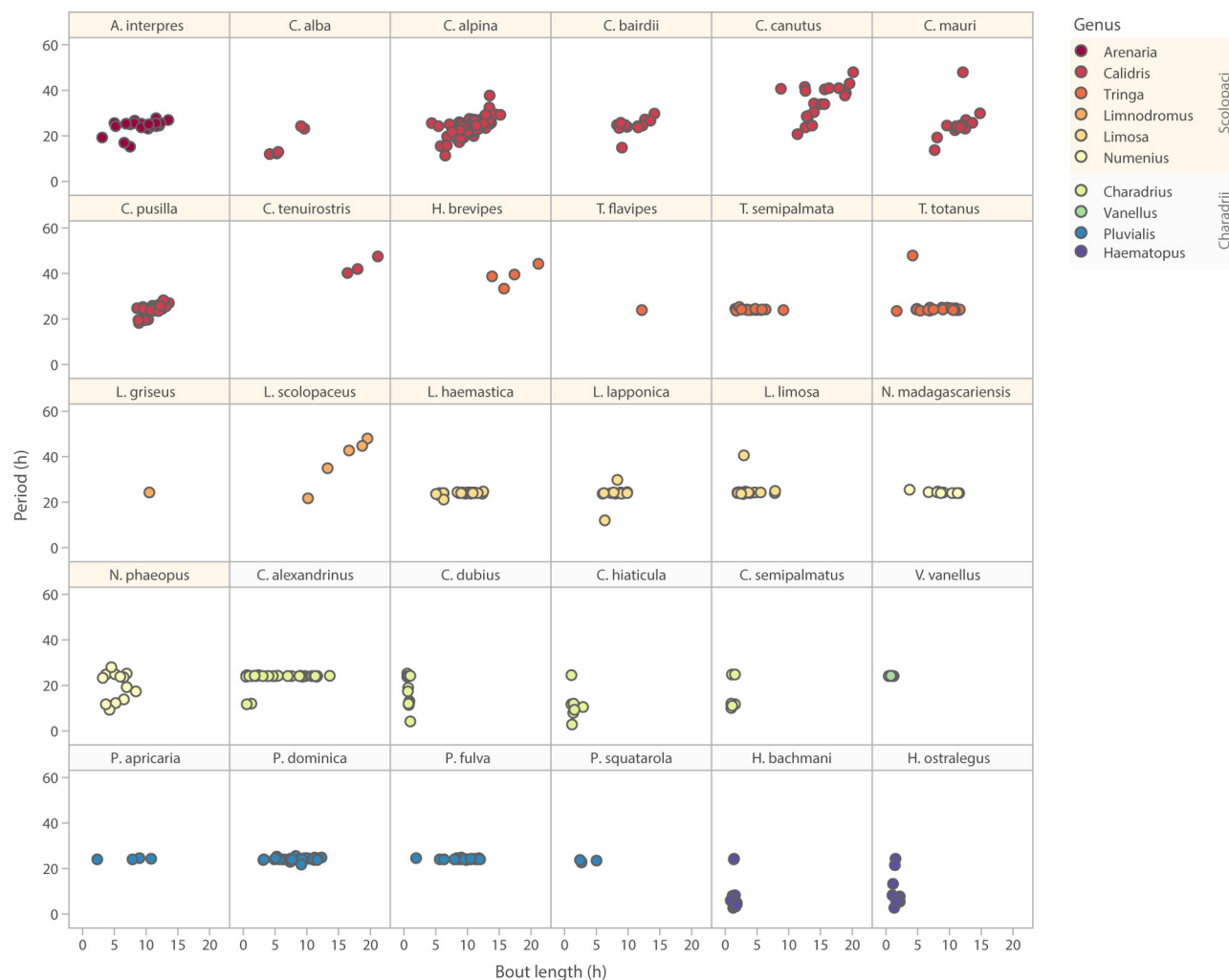
Extended Data Figure 1 | Extracting period length of incubation rhythms. **a–c,** Each column represents an example for a specific nest with long, intermediate and short incubation bouts. **a,** From the extracted bout lengths we created a time series that indicated—for each nest and for every 10 min interval—whether a specific parent (female, if sex was known) incubated or not. Exchange gaps (no parent on the nest) had to be <6 h to be included (for treatment of exchange gaps >6 h see **d, e**). **b,** We then estimated the autocorrelation for each 10-min time-lag up to 4 days (R ‘acf’ function⁶³). Positive values indicate a high probability that the female was incubating, negative values indicate that it was more likely that the male was incubating. We used only nests that had enough data to estimate the autocorrelation pattern ($n = 584$ nests from 88 populations of 30 species). The visualized autocorrelation time series never resembled white or random noise indicative of an arrhythmic incubation pattern. To determine the period (that is, cycle of high and low probability for a parent to incubate) that dominated the incubation rhythm, we fitted to the autocorrelation estimates a series of periodic logistic regressions. In each regression, the time lag (in hours) transformed to radians was represented by a sine and cosine function $f(t) = a_0 + b\left(\cos \frac{2\pi t}{T}\right) + c\left(\sin \frac{2\pi t}{T}\right) + e$, where $f(t)$ is the autocorrelation at time-lag t ; a_0 is the intercept; b is the

slope for sine and c the slope for cosine, T represents the length of the fitted period (in hours), and e is an error term. We allowed the period length to vary from 0.5 h to 48 h (in 15 min intervals, giving 191 regressions). **c,** By comparing the Akaike’s information criterion⁶⁴ (AIC) of all regressions, we estimated, for each nest, the length of the dominant period in the actual incubation data (best fit). Regressions with ΔAIC ($\text{AIC}_{\text{model}} - \text{AIC}_{\text{min}}$) close to 0 are considered as having strong empirical support, while models with ΔAIC values ranging from 4–7 have less support⁶⁴. In 73% of all nests, we determined a single best model with $\Delta\text{AIC} \leq 3$ (**c**, middle ΔAIC graph), in 20% of nests two best models emerged and in 6% of nests 3 or 4 models had $\Delta\text{AIC} \leq 3$ (**c**, left and right ΔAIC graphs). However, in all but three nests, the models with the second-, third- and so on best ΔAIC were those with period lengths closest to the period length of the best model (**c**, left and right ΔAIC graphs). This suggests that multiple periodicities are uncommon. **d, e,** The extraction of the period length (described in **a–c**) requires continuous data sets, but some nests had long (>6 h) gaps between two consecutive incubation bouts, for example because of equipment failure or because of unusual parental behaviour. In such cases, we excluded the data from the end of the last bout until the same time the following day, if data were then available again (**d**), or we excluded the entire day (**e**).

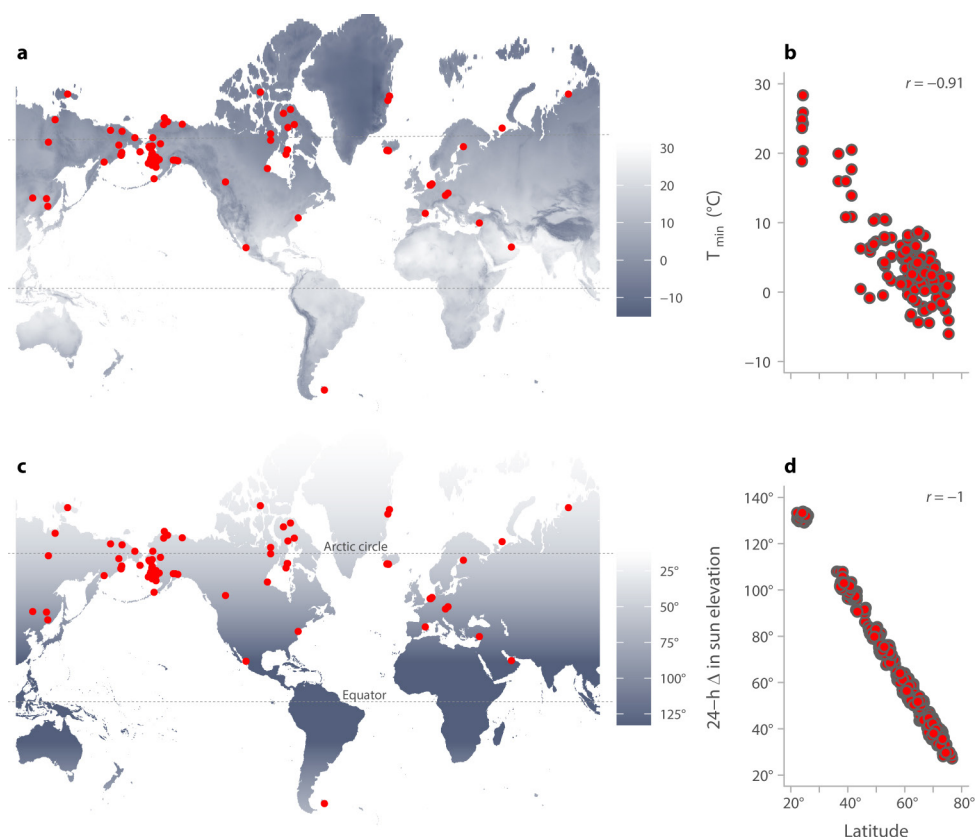


Extended Data Figure 2 | Extracting incubation bouts from light-logger data. **a**, An example of a nest with a light intensity signal from both parents (yellow line, female; blue line, male). The incubation bouts for a given parent reflect periods dominated by lower light values compared to those of the partner. Note the sharp drop in the light levels at the beginning of each incubation bout and the sharp increase in the light levels at the end. Change-overs between partners occur when the light signal lines cross. Such pronounced changes in light intensity detected by the logger were used to assign incubation even when only a single parent was tagged. Note that after the chicks hatch and leave the nest (9 July, vertical bar), the light intensity signals from both parents remain similar. **b**, An example

of a nest where one incubating parent was simultaneously equipped with a light-logger and with a GPS tag. The yellow line indicates light levels, red dots indicate the distance of the bird to the nest. As expected, low light levels co-occur with close proximity to the nest, and therefore reflect periods of incubation. Although light levels decrease during twilight (light grey horizontal bar), the recordings were still sensitive enough to reflect periods of incubation, that is, the light signal matches the distance (for example 25 May: female incubated during dawn, but was off the nest during dusk). **a**, **b**, Rectangles in the background indicate incubation bouts (female, light yellow polygon; male, light blue polygon).



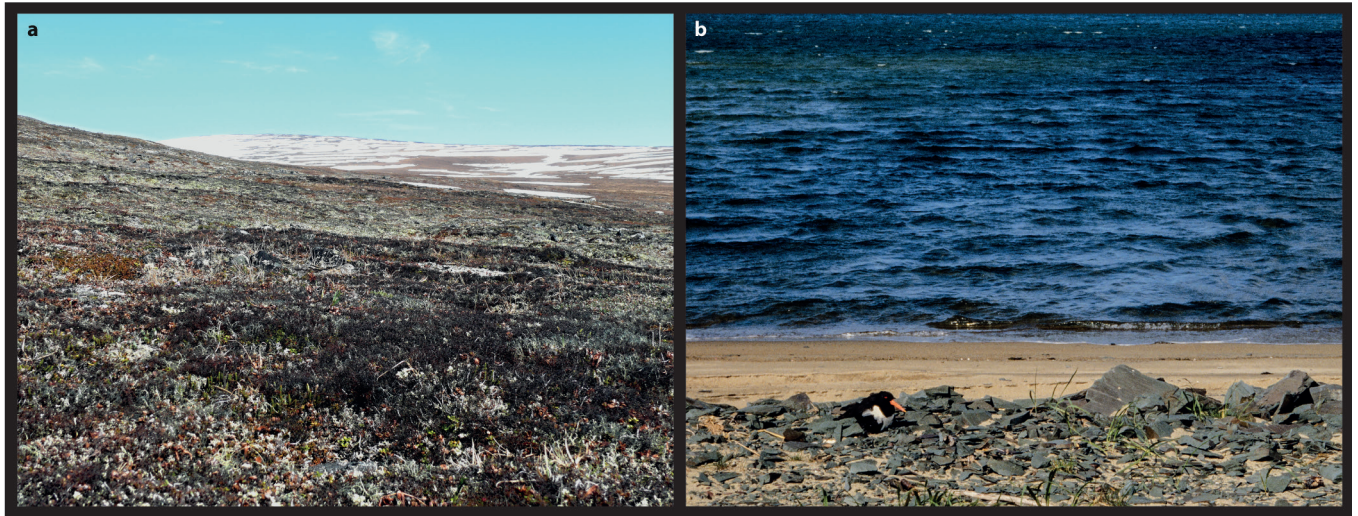
Extended Data Figure 3 | Relationship between bout and period length for 30 shorebird species. Each dot represents one nest ($n = 584$ nests), colours indicate the genus.



Extended Data Figure 4 | Ecological correlates of latitude. **a**, Variation in minimum temperature across the globe represented by mean minimum June temperature for the Northern Hemisphere and mean minimum December temperature for the Southern Hemisphere. **b**, Correlation between absolute latitude and the mean minimum temperature of the month ($n = 729$ nests). For each nest we used the month that contained most of the incubation data. For maximum temperature the correlation was the same ($r = -0.91$, $n = 729$ nests). **c**, Daily variation in sun elevation (that is, in light conditions) are represented as the difference between the

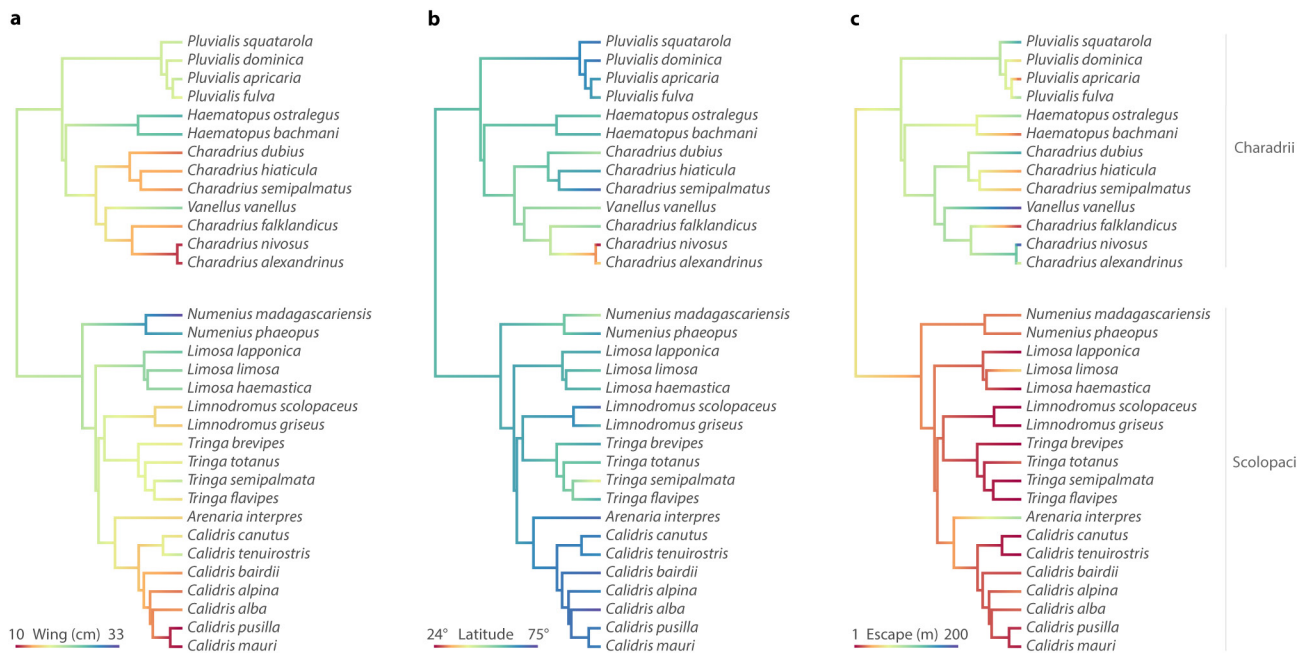
noon and midnight sun elevation for the summer solstice in the Northern Hemisphere and the winter solstice in the Southern Hemisphere.

d, Correlation between absolute latitude and daily variation in sun elevation for mid-day of incubation data for each nest ($n = 729$ nests). The points are jittered, as otherwise they form a straight line. **a**, **c**, Red points indicate the breeding site for each population ($n = 91$). **a**, **b**, The minimum and maximum monthly temperature data were obtained from <http://www.worldclim.org> using the raster R-package⁶⁵. **c**, **d**, Sun-elevation was obtained by the 'solarpos' function from the maptools R-package⁶⁶.



Extended Data Figure 5 | Between-species variation in parental crypsis during incubation. **a, b,** Shorebirds vary in how visible they are on the nest while incubating. The nearly invisible great knot (*Calidris tenuirostris*; **a**; central and facing right) sits tight on the nest when approached by a

human until nearly stepped upon. In contrast, the conspicuous Eurasian oystercatcher (*Haematopus ostralegus*; **b**) is visible on the nest from afar and when approached by a human leaves the nest about 100 m in advance (Credits: **a**, M. Šálek; **b**, J. van de Kam).



Extended Data Figure 6 | Phylogenetic relationships for predictors. **a**, Body size, estimated as female wing length. **b**, Latitude (absolute). **c**, Escape distance. **a–c**, We visualized the evolution of these traits^{29,67} using the median (**a**, **b**; based on population medians), estimates of escape distance for each species (**c**) and one of the 100 sampled trees (see Methods).

Extended Data Table 1 | Incubation monitoring methods and systems

Method	Model	Company	www	Year	Sampling interval (min)	<i>n</i> populations	<i>n</i> nests	<i>n</i> both parents tagged
RFID + tag embedded in the flag ^{31,32}	tag: 9.0 × 2.1 mm, 0.087 g, 134.2 kHz	Biomark	http://www.biomark.com/	2011-2014	0.08	19	200	157
	RFID: custom made	Calima Engineering, & Max Planck Institute for Ornithology	http://www.calima.de http://www.orn.mpg.de/en					
	temperature probe: TinyTag Talk PB-5005-0M6 (ø 2.5 mm) with Talk 2 logger TK-4023 or	Gemini Data Loggers Ltd	www.tinytag.info					
	MSR® (12 × 6 mm) with MSR® 145 logger	MSR® Electronics GmbH	http://www.msr.ch/					
	tag: 11.5 × 2.12 mm, 0.1 g, 125 kHz	Trovan®	http://www.trovan.com	2007	5	1	3	3
	RFID: custom made	EID Aalten BV	http://www.dorset.nu/en/					
RFID + tag glued to the tail ³³	temperature probe: TinyTag Talk PB-5005-0M6 (ø 2.5 mm) with Talk 2 logger TK-4023	Gemini Data Loggers Ltd	www.tinytag.info					
	tag: 11.5 × 2.12 mm, 0.09g, 128 kHz	Trovan®	http://www.trovan.com	2005-2008	1.7-5.5	1	34	34
	RFID: LID650 with extended memory, driver TM613, Antenna ANT614							
	tag: TIRIS 12 × 6 × 3 mm, 0.4 g, 134.2 kHz	Texas Instruments	http://www.ti.com	1997	1/3	1	6	6
	RFID: TIRIS Micro-reader							
	TIRIS tag: 12 × 6 × 3 mm, 0.4 g, 134.2 kHz	Texas Instruments	http://www.ti.com	2006	1/3	1	18	18
Light logger ^{34,35,68}	RFID: custom made	Francis Scientific Instruments Limited	https://www.duedil.com/comp-any/01964877/francis-scientific-instruments-limited					
	MK10, 12, 14 18, 20: 0.7-1.4g, light scale 0-64 (64 ≥ 100 lux)	British Antarctic Survey, currently Biotrack	http://www.biomark.com/	2008-2013	2-10	48	261*	55
	Intigeo W65A9RK: 0.7g, absolute light levels (lux)	Migrate Technology Ltd	http://www.migratetech.co.uk	2012-2014	5-10	23	136*	11
GPS-tracker ³⁶	UvA-BITS 4C: 62 × 30 × 12, 14 g	University of Amsterdam	http://www.uva-bits.nl/	2010	10-30	1	8	8
	UvA-BITS 2CDse: 52 × 22 × 9, 7.5 g			2014	20	1	1**	0
Radio-transmitter ²⁴	iTag: 26 × 15 × 9 mm, 4 g	e-obs GmbH	http://www.e-obs.de/	2009	0.07	2	3	3
Video ⁶⁹⁻⁷²				2005-2006	1/3	1	28	-
				2005-2011	constant	3	25	-
Observations				1994	30	1	5	-
				2011	constant	1	3	-

For details about methods used in each populations, see supplementary data of ref. 53.

*At one nest a bird with a MK logger was recaptured and the logger exchanged for an Intigeo logger. This nest appears in *n* for both logger types.

**Simultaneously equipped with light-logger (Intigeo). This nest appears in *n* for both GPS-tracker and Intigeo.

Extended Data Table 2 | Effects of phylogeny and sampling on bout length and period length

Response	Effect type	Effect	Posterior mode	95% CI		N (range)
				Lower	Upper	
Median bout [h]	Fixed	Intercept	7.2	1.04	12	1100 (924-2079)
		Sampling	0.16	-0.2	0.61	1100 (809-1644)
	Random (variance)	Phylogeny	25.33	4.6	59.6	1100 (753-1383)
		Species	0.01	0	12.1	1100 (779-1636)
		Breeding site	2.13	0.96	4.28	1100 (808-2242)
		Residual	5.04	4.51	5.61	1100 (838-1444)
	Page's λ		1	0.5	1	1100 (814-1316)
Period [h]	Fixed	Intercept	21.94	12.8	30.67	1100 (765-1392)
		Sampling	0.13	-0.41	0.65	1100 (741-1468)
	Random (variance)	Phylogeny	66.22	14.3	153	1100 (729-1638)
		Species	0.06	0	29.36	1100 (729-1435)
		Breeding site	0.01	0	0.88	1100 (814-1378)
		Residual	14.87	13.3	16.84	1100 (884-1460)
	Page's λ		1	0.54	1	1100 (740-1523)

The posterior estimates (modes) of the effect sizes with the highest posterior density intervals (95% CI) and the median and range of the effective sample sizes (N (range)) come from the joint posterior distribution of 100 separate runs each with 1 of 100 separate phylogenetic trees from <http://birdtree.org>. $n_{\text{bout}} = 729$ nests from 91 populations belonging to 32 species. $n_{\text{period}} = 584$ nests from 88 populations belonging to 30 species. Sampling (how often the incubation behaviour was sampled) was ln-transformed and then mean-centred and scaled (divided by s.d.). For procedures and specifications related to phylogenetic Bayesian mixed models see Methods. Estimating Page's λ on the species level ($n_{\text{bout}} = 32$ species, $n_{\text{period}} = 30$ species) with phylogenetic generalized least-squares using the function 'pgls' from the R package *caper*⁷³ gave similar results (median (range) $\lambda_{\text{bout}} = 0.73$ (0.63–1) and $\lambda_{\text{period}} = 0.95$ (0.64–1), based on 100 estimates each for 1 of the 100 trees).

Extended Data Table 3 | Source of phylogenetic signal

Suborder included	Response	Effect type	Fixed effects	Posterior mode	95% CI		N
					Lower	Upper	
No	Bout [h]	Fixed	Intercept	7.69	4.11	10.76	1100
		Random (variance)	Genus	9.54	0.02	46.6	1100
			Species	11	4.84	25.77	1100
			Breeding site	2.25	1.06	4.5	1100
			Residual	5.03	4.55	5.61	1100
Yes	Bout [h]	Fixed	Intercept (Charadrii)	4.48	1.05	7.66	1100
			Suborder (Scolopaci)	6.07	1.52	10.44	1100
		Random (variance)	Genus	0.09	0	21.35	1100
			Species	11.29	5.53	23.99	1275
			Breeding site	2.31	0.98	4.28	1100
			Residual	5.04	4.52	5.58	1100
No	Period [h]	Fixed	Intercept	23.46	18.22	27.39	1100
		Random (variance)	Genus	19.05	0	97.79	1100
			Species	41.02	17.04	72.66	1100
			Breeding site	0.01	0	0.9	1100
			Residual	14.94	13.35	16.76	1100
Yes	Period [h]	Fixed	Intercept (Charadrii)	18.61	13.12	23.98	940
			Suborder (Scolopaci)	8.92	0.6	15.4	990
		Random (variance)	Genus	0.33	0	59.09	769
			Species	34.41	16.91	68.14	1142
			Breeding site	0.01	0	0.84	1100
			Residual	15.01	13.34	16.86	891

The posterior estimates (modes) of the effect sizes with the highest posterior density intervals (95% CI) and the effective sample sizes (N) come from a posterior distribution of 1,100 simulated values generated by MCMCglmm in R⁶². $n_{\text{bout}} = 729$ nests from 91 populations belonging to 32 species. $n_{\text{period}} = 584$ nest from 88 populations belonging to 30 species.

Extended Data Table 4 | Effect of latitude, body size, escape distance and life history on biparental incubation rhythms in shorebirds

Response	Effect type	Fixed effects	Posterior mode	95% CI		N (range)
				Lower	Upper	
Bout [h]	Fixed	Intercept	7.45	2.65	12	1100 (804-1496)
		Wing length	-0.78	-2.5	1.05	1100 (839-1638)
		Latitude	1.72	0.63	2.65	1100 (850-1642)
		Escape distance	-1.68	-3.3	-0.25	1100 (634-2046)
	Random (variance)	Phylogeny	0.19	0	45	1100 (803-1875)
		Species	0.07	0	14.4	1100 (695-1580)
		Breeding site	1.4	0.59	3.02	1100 (833-1480)
		Residual	5.02	4.53	5.64	1100 (516-1916)
	Page's λ		0.72	0.13	1	1100 (731-1407)
Light entrainable rhythm [1,0] on binomial scale	Fixed	Intercept	-1.62	-3.19	-0.13	1100 (731-1633)
		Latitude	-0.56	-1.15	-0.07	1100 (765-1575)
	Random (variance)	Phylogeny	0.05	0	5.54	1100 (883-1371)
		Species	0.02	0	2.68	1100 (965-2246)
		Breeding site	0	0	0.63	1100 (605-1304)
	Page's λ		0.74	0.02	1	1100 (932-1498)
Absolute deviations from 24-h	Fixed	Intercept	0.17	-0	0.35	1100 (459-1501)
		Latitude	0.03	-0	0.07	1100 (777-1488)
	Random (variance)	Phylogeny	0	0	0.07	1100 (786-1393)
		Species	0	0	0.03	1100 (861-1412)
		Breeding site	0	0	0	1100 (826-1860)
		Residual	0.03	0.03	0.04	1100 (948-2039)
	Page's λ		0.74	0.02	1	1100 (843-1471)
Deviations from 24-h	Fixed	Intercept (non-tidal)	0.02	-0.04	0.09	1100 (851-1742)
		Life history (tidal)	-0.02	-0.1	0.04	1100 (702-2257)
	Random (variance)	Phylogeny	0	0	0.01	1100 (806-1692)
		Species	0	0	0	1100 (692-1601)
		Breeding site	0	0	0.01	1100 (656-1490)
		Residual	0.07	0.06	0.08	1100 (760-1563)
	Page's λ		0.77	0.01	1	1100 (864-1451)

The posterior estimates (modes) of the effect sizes with the highest posterior density intervals (95% CI) and the median and range of the effective sample sizes (N (range)) come from the joint posterior distribution of 100 separate runs each with 1 of the 100 separate phylogenetic trees from <http://birdtree.org>. $n_{\text{bout}} = 729$ nests from 91 populations belonging to 32 species. For models on light-entrainable rhythm, absolute deviations and deviations from 24 h: $n = 584$ nests from 88 populations belonging to 30 species. Latitude (in bout model: absolute value), wing length (ln-transformed), and escape distance (ln-transformed) were mean-centred and scaled (divided by s.d.). The estimates for the light-entrainable rhythm are on a binomial scale. For procedures and specifications related to phylogenetic Bayesian mixed models see Methods.

Genomic evolution and chemoresistance in germ-cell tumours

Amaro Taylor–Weiner^{1,2*}, Travis Zack^{1,3*}, Elizabeth O'Donnell^{4,5}, Jennifer L. Guerriero⁴, Brandon Bernard⁴, Anita Reddy⁶, G. Celine Han^{2,4}, Saud AlDubayan^{7,8}, Ali Amin–Mansour², Steven E. Schumacher^{2,9}, Kevin Litchfield^{10,11}, Clare Turnbull^{10,11}, Stacey Gabriel², Rameen Beroukhi^{2,4}, Gad Getz^{2,12}, Scott L. Carter^{2,13,14,15}, Michelle S. Hirsch¹⁶, Anthony Letai⁴, Christopher Sweeney^{4§} & Eliezer M. Van Allen^{2,4,13§}

Germ-cell tumours (GCTs) are derived from germ cells and occur most frequently in the testes^{1,2}. GCTs are histologically heterogeneous and distinctly curable with chemotherapy³. Gains of chromosome arm 12p and aneuploidy are nearly universal in GCTs^{4–6}, but specific somatic genomic features driving tumour initiation, chemosensitivity and progression are incompletely characterized. Here, using clinical whole-exome and transcriptome sequencing of precursor, primary (testicular and mediastinal) and chemoresistant metastatic human GCTs, we show that the primary somatic feature of GCTs is highly recurrent chromosome arm level amplifications and reciprocal deletions (reciprocal loss of heterozygosity), variations that are significantly enriched in GCTs compared to 19 other cancer types. These tumours also acquire *KRAS* mutations during the development from precursor to primary disease, and primary testicular GCTs (TGCTs) are uniformly wild type for *TP53*. In addition, by functional measurement of apoptotic signalling (BH3 profiling) of fresh tumour and adjacent tissue⁷, we find that primary TGCTs have high mitochondrial priming that facilitates chemotherapy-induced apoptosis. Finally, by phylogenetic analysis of serial TGCTs that emerge with chemotherapy resistance, we show how TGCTs gain additional reciprocal loss of heterozygosity and that this is associated with loss of pluripotency markers (*NANOG* and *POU5F1*)^{8,9} in chemoresistant teratomas or transformed carcinomas. Our results demonstrate the distinct genomic features underlying the origins of this disease and associated with the chemosensitivity phenotype, as well as the rare progression to chemoresistance. These results identify the convergence of cancer genomics, mitochondrial priming and GCT evolution, and may provide insights into chemosensitivity and resistance in other cancers.

GCTs arise from germ cells, which can be found in the gonads (testicles and ovaries), mediastinum, pituitary and retroperitoneum, and progress from precursor lesions (germ-cell neoplasia *in situ*; GCNIS)^{1,2,10}. Histologically, GCTs are classified as seminoma or non-seminoma. The latter may consist of one histology or be mixed non-seminoma. Clinically, most GCTs are exceptionally chemosensitive; combination cisplatin- and etoposide-based chemotherapy cures more than 80% of GCT patients with metastatic disease³. However, approximately 10% of patients with metastatic GCTs will die as a result of chemoresistant disease.

Leveraging preclinical models, several mechanisms have been proposed for GCT oncogenesis and chemosensitivity, including DNA

repair deficiencies, embryonic stem (ES) cell-like cell cycle properties, or high apoptotic propensity^{11–13}. Germline genome wide association studies have identified several risk loci for GCT, most notably in *KITLG*^{14,15}. For GCTs, karyotyping demonstrated marked aneuploidy⁶, targeted sequencing identified rare *KRAS* hotspot mutations¹⁶, and whole-exome sequencing (WES) confirmed nearly universal arm-level chromosome arm 12p gain⁴, low mutation rates, and activating *KIT* mutations specific to seminomas⁵. However, genomic features in GCTs that contribute to the origins, chemosensitivity phenotype, and progression remain incompletely characterized.

Comparison of serial treatment-responsive and refractory tumours has informed genomic progression in other cancers^{17,18}. We proposed that genomic evolutionary analysis of clinical GCTs from patients who respond and become resistant to chemotherapy would identify features underlying disease progression and chemotherapy response phenotype. We performed clinically integrated molecular analysis of 59 tumours samples (with matched germ line) from 47 patients with TGCTs and two patients with primary mediastinal GCTs (PMGCTs) to identify genomic features associated with disease origin and progression. The aggregate cohort features are summarized in Fig. 1a and Supplementary Tables 1, 2.

The median mutational load in our cohort was 0.9 mutations Mb^{–1} (Supplementary Table 1). Mutational significance analysis¹⁹ identified *KRAS* as the most statistically significant altered gene in this cohort (Fig. 1a, b, Supplementary Table 3). The only other significantly mutated gene was *RPL5*, a ribosomal protein previously implicated in other cancer types¹⁹ (Fig. 1b). *RPL5* may regulate the MDM2–TP53 axis²⁰ but was mutated infrequently and is of unknown biological relevance in GCTs. *KIT* mutations were not observed in our cohort despite sufficient genomic power to detect mutations (*KIT* hotspot loci mean coverage: 316×; range: 89–800×), although meta-analysis of our cohort with a separate TGCT WES set⁵ confirms *KIT* as significantly mutated (Extended Data Fig. 1a). Expanded investigation for clinically or biologically relevant alterations identified singleton mutations in DNA repair or sex determination pathway gene sets (see Supplementary Table 4 for all somatic mutations), although these were rare and of unclear significance.

Because *KRAS* emerged as the primary significantly mutated gene, we next sought to explore the point of *KRAS* mutation emergence in the context of TGCT progression. We performed phylogenetic analysis of WES from patient-matched GCNIS tumours (pre-invasive) and

¹Division of Medical Sciences, Harvard University, Boston, Massachusetts 02115, USA. ²Cancer Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ³Health Sciences and Technology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ⁵Department of Medical Oncology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁶Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁷Division of Genetics and Genomics, Department of Medicine, Boston Children's Hospital, Massachusetts 02115, USA. ⁸Department of Medicine, King Saud bin Abdulaziz University for Health Sciences, Saudi Arabia. ⁹Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ¹⁰Division of Genetics and Epidemiology, The Institute of Cancer Research, Fulham Road, London SW3 6JB, UK. ¹¹William Harvey Research Institute, Queen Mary University London, Charterhouse Square, London EC1M 6BQ, UK. ¹²Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ¹³Center for Cancer Precision Medicine, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ¹⁴Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. ¹⁵Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA. ¹⁶Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

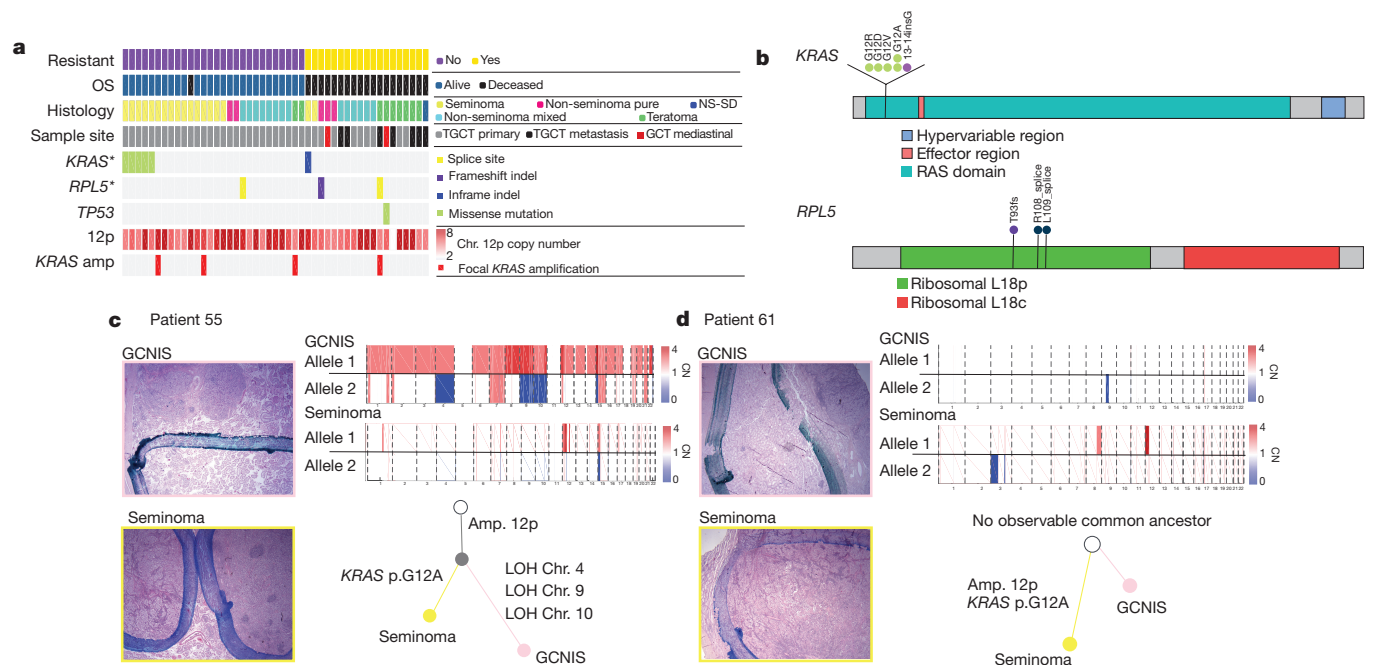


Figure 1 | Mutational landscape and evolution from precursor lesions.

a, Clinical and genomic overview. Each column represents a patient sample. The top four rows represent chemoresistant disease status, overall survival (OS) vital status, predominant histological classification of the primary sample (when available, or selected secondary sample otherwise), and anatomic site, respectively. The next three rows indicate the patient's mutation status for the two significantly ($q < 0.1$) mutated genes, and *TP53*. Next median copy number of chromosome arm 12p is shown, followed by the focal amplification status of *KRAS* (defined as *KRAS* copy number greater than $2 \times$ median 12p copy number). NS-SD refers to non-seminoma with sarcoma differentiation. **b**, Mutations in *KRAS* and

RPL5. Each circle represents a mutation observed in a single patient, and labels indicate amino acid change. **c**, **d**, Phylogenetic relationship between GCNIS and primary TGCT. Histological slide images from DFCL_55 and DFCL_61 are shown on the left. Images in the pink borders are GCNIS and below in yellow are tumour images. Dark blue ink indicates the border of the lesion. Next to the histological images are phylogenetic trees determined by shared allelic copy number (CN) mutations. The non-shaded circle at the top of the tree indicates the patient's germ line. Below that the filled grey is the inferred shared ancestor clone and then each sample branches ends in the sequenced sample. Above are allelic copy number data supporting the tree.

primary TGCT tumours (Methods). In one representative case, both samples contained chromosome arm 12p gain; however, the only somatic putative driver mutation distinguishing the two samples was a *KRAS*^{G12A} mutation in the TGCT (0 out of 377 reads in GCNIS, 65 out of 340 reads in TGCT) (Fig. 1c, Extended Data Fig. 2a). A similar pattern of *KRAS* mutation exclusive to the TGCT and not in the patient-matched GCNIS was also seen in a second patient (0 out of 210 reads in GCNIS, 35 out of 184 reads in TGCT) (Fig. 1d, Extended Data Fig. 2b). Although not required for progression, activating *KRAS* mutations may occur after arm level gain of chromosome 12p, thereby indicating that these are two separate processes involved in TGCT evolution.

While GCTs had low point mutation rates, nearly all tumours contained arm level gain of chromosome arm 12p, as previously described⁵. Given previous reports of aneuploidy in GCTs^{6,21}, we performed allelic and absolute copy number analysis through allele deconstruction^{22,23} to determine whether there were specific copy number patterns in GCTs not been previously identifiable or reported (Methods). In a representative case (DFCL_7; Fig. 2a), coverage-based copy number analysis identified 12p gain, but only allelic copy number analysis revealed frequent arm and chromosome level gains of one parental allele with simultaneous loss of the other parental allele, leading to loss of heterozygosity (LOH). These reciprocal LOH (RLOH) events often maintained the germline number of DNA copies (copy-neutral LOH), but also frequently have further amplification of the remaining parental allele. For example, in DFCL_7 chromosome 4 has undergone copy-neutral LOH, whereas chromosome 3q contains a LOH event in which the total copy number is three. Across the GCT cohort, the mean number of chromosomal arm level amplifications was 28.3 ± 7.5 (mean \pm s.d.), which is significantly increased compared to 6,509 tumours representing 19 other cancer types ($P < 0.0001$; Mann-Whitney) (Fig. 2b, Extended Data Fig. 3 and Supplementary Tables 5, 6).

Approximately 45% of arm level deletions contained a compensatory reciprocal amplification in GCTs, again sharply increased compared to other tumour types ($P < 0.0001$; Mann-Whitney) (Fig. 2c).

To validate this RLOH pattern, we performed the same analysis on a separate primary TGCT fresh-frozen WES cohort (ICR GCT)⁵ and observed similar frequencies of reciprocal copy number alterations (Fig. 2b, c, Extended Data Fig. 1b–d). Given the prevalence of arm level amplifications observed in GCTs, we performed a permutation test involving deleted chromosomal arms to examine whether these deletions would be compensatory by chance²³ (Methods). We observed significantly increased reciprocal events in GCTs than expected by chance (empirical $P < 0.0001$) (Fig. 2d). Finally, to examine whether this copy number feature was an intrinsic germ-cell property and not related to microenvironmental features, we performed absolute copy number analysis on three tumours originating in the testes with non-germ-cell histologies and did not identify such chromosomal abnormalities (Extended Data Fig. 4). Overall, GCTs have a unique pattern of highly recurrent reciprocal copy number alterations in both discovery and validation cohorts.

Despite extensive copy number events in GCTs, all TGCTs were *TP53* wild type, and we only observed one *TP53* mutation in a PMGCT. Transcriptome profiling of a patient subset ($n = 28$: 21 primary TGCTs, five metastatic samples from TGCTs, and two PMGCT; Methods) demonstrated *TP53* expressed across histologies (Fig. 3a, Supplementary Table 7). The overall constellation of genomic changes in clinical TGCT tumours—wild-type *TP53*, chromosome arm 12p gain, and additional recurrent copy number changes—is similar to adaptive mutations acquired by human ES cells during prolonged passaging *in vitro*^{24,25}. Like GCTs, ES cells are acutely sensitive to DNA damage induced by chemotherapy, and it has been shown that this sensitivity is a result of intact p53 status and high mitochondrial priming²⁶.

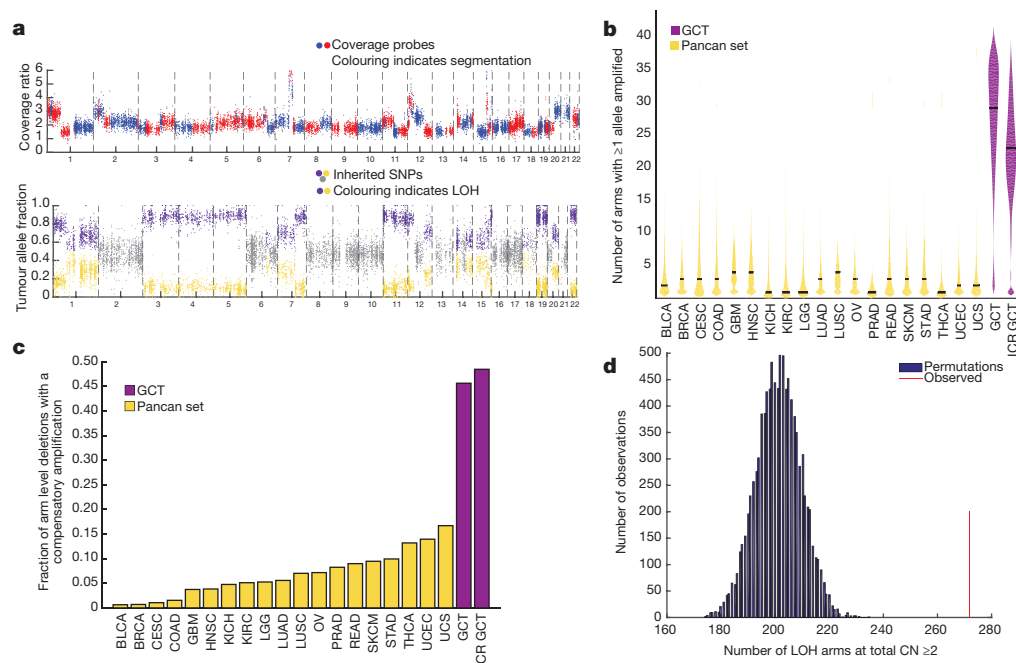


Figure 2 | RLOH in GCTs. **a**, Representative allelic somatic copy number data. Top, coverage-derived copy number events in a tumour. Bottom, the allele fraction of germ-line variants observed in the tumour. Regions in which there is LOH are coloured in purple and orange for the retained and lost allele, respectively. **b**, Arm level amplification rates in GCTs and 19 other tumour types (see <https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations> for abbreviations). Shown are the distributions of counts of arm level amplifications, after controlling for whole-genome doubling, in samples from each tumour type. The distributions are estimated using a Gaussian kernel density³¹ and the median count is indicated by the black bar. GCTs (purple, with the independent replication series denoted as ICR GCT) have significantly more arm level amplifications than other tumour types (gold) ($P < 0.0001$;

Mann–Whitney). **c**, Rate of reciprocal deletion on arms with amplification in GCTs and 19 other tumour types. For each sample, we counted the number of arm level amplifications with and without a reciprocal deletion. Shown are the mean fractions of amplified arms with a reciprocal arm level deletion, after controlling for whole-genome doubling. GCTs (purple) have significantly more reciprocal deletions than other tumour types (gold) ($P < 0.0001$; Mann–Whitney). **d**, Permutation test of arm level deletions. To generate a null distribution of reciprocal deletions, we shuffled arm level deletions within samples²³ and for each permutation we counted the number of observed reciprocal arm level deletions. Shown is a histogram of the counts of reciprocal deletions observed in permutations. The red line indicates the number observed in our dataset corresponding to an empirical $P < 0.0001$.

Also, one of the earliest identified susceptibility loci in TGCTs is in *BAK1*, a *BCL2* antagonist and pro-apoptotic gene¹⁴, and chemosensitivity in other cancer types has been associated with increased mitochondrial priming⁷. We therefore proposed that the basis of chemosensitivity in TGCTs with a wild-type *TP53* and RLOH abundant genomic background was a result of a fundamental apoptotic propensity caused by increased mitochondrial priming. Mitochondrial priming refers to the intrinsic potential of cellular apoptotic propensity due to the balance of pro-apoptotic and anti-apoptotic *BCL2* family proteins at the mitochondria, a feature that can be assessed by BH3 profiling²⁷. To test this hypothesis, we obtained fresh tumour samples from seven TGCTs (Fig. 3b) and performed dynamic BH3 profiling on the tumours and adjacent normal tissue⁷ (Methods). In one representative case, we observed an increase in BIM BH3-induced mitochondrial depolarization between the tumour and adjacent normal (Fig. 3c) ($P < 0.0001$; paired *t*-test). Across the seven samples, BIM BH3-induced mitochondrial priming remained significant (Fig. 3d) ($P < 0.0001$; paired *t*-test). Thus, intact *TP53*, RLOH, and high mitochondrial priming may form the foundation of chemosensitivity in TGCTs.

Although most TGCTs are cured with chemotherapy, some recur and approximately 10% of patients with metastatic disease will die from chemoresistant TGCT. To track the genomic evolution of chemoresistant TGCTs, phylogenetic analysis of 13 WES tumour samples (pre- and post-chemotherapy) from 5 patients was performed (Fig. 4, Extended Data Fig. 5, Supplementary Table 8, Methods). Chromosome arm 12p was a truncal event shared in all samples. In each patient, the resistant tumour that arose after either first- or second-line chemotherapy accumulated additional copy number events, including reciprocal deletions and events involving new chromosomes (Fig. 4). In one

patient with serial tumours resected over a 13-month period without intervening chemotherapy, the copy-number-derived phylogeny was proportional to the sequence and elapsed time of sample acquisition (Fig. 4a). In two cases in which the final post-chemotherapy sample had teratoma and/or poorly differentiated carcinoma components consistent with differentiation (Fig. 4b, c), complete genome doubling was observed. Notably, the pluripotency and apoptosis regulators *NANOG* and *POU5F1* (also known as *OCT3/4*) expressed in GCT^{8,9,13} were not expressed in the tumours obtained from metastatic deposits resistant to chemotherapy (Fig. 4e; $P < 0.001$, Mann–Whitney). They were also not expressed in the two incurable PMGCTs (Fig. 4e, Supplementary Table 7). Taken together, chemotherapy-resistant germ-cell tumours are associated with continued progression of RLOH copy number events and loss of pluripotency markers in this clinical cohort, although it is uncertain whether the loss of pluripotency markers is a driver of chemoresistance^{11,28}, a marker of methylation changes in a further differentiated histology post-chemotherapy²⁹, or a feature of both processes.

Overall, while GCTs represent a heterogeneous set of histological subtypes, they share a distinct burden and pattern of reciprocal copy number alterations. This observation is consistent with the notion that GCTs emerge, in part, through dysregulation in the mitosis/meiosis switch from arrested gonocytes, since this process may result in abnormal chromosomal segregation³⁰ and persist during the evolution of GCTs. Expanded molecular analysis may inform the initiators of RLOH and its role in oncogenesis. Subsets of these tumours acquire *KRAS*-activating mutations, although the necessity for *KRAS* mutation is less clearly defined, and these tumours lack recurrent nonsynonymous mutations in established cancer genes¹⁹. Instead, chemosensitivity in TGCTs may be the result of high mitochondrial priming properties, and

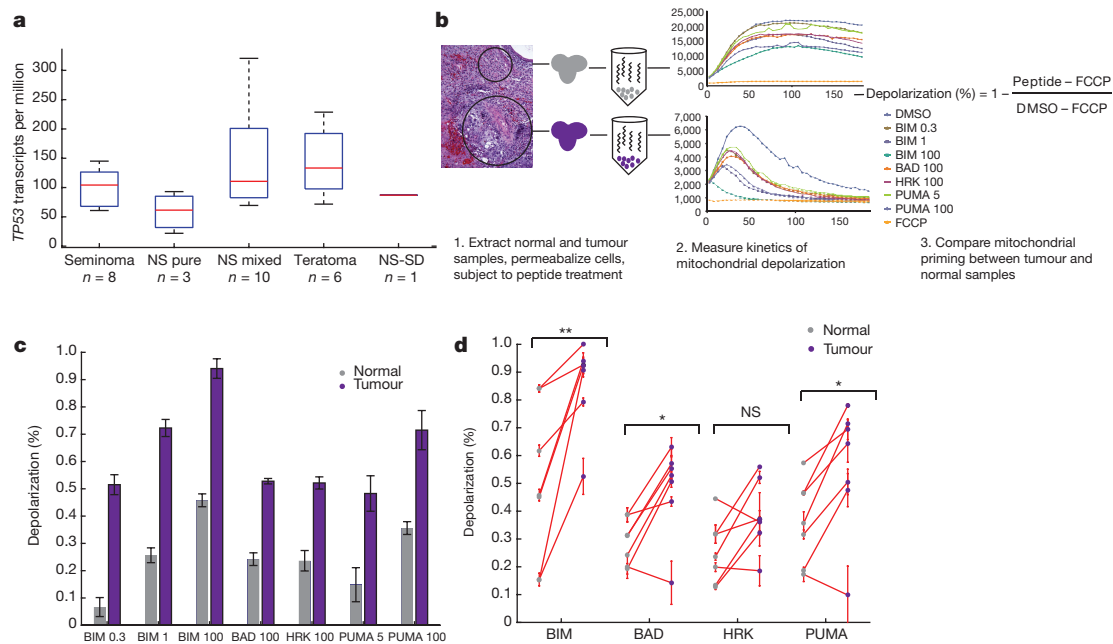


Figure 3 | Mitochondrial priming in germ cell tumours. **a**, Gene expression of *TP53* in GCTs. Values are in transcripts per million (TPM). **b**, BH3 profiling pipeline schematic. Samples were collected surgically and then processed immediately as previously described⁷. Mitochondrial priming was measured in both the tumour and normal samples after exposure to each peptide at varying concentrations. The priming caused by each peptide is compared to the negative and positive controls (dimethylsulfoxide (DMSO) and the uncoupler carbonyl cyanide-*p*-trifluoromethoxyphenylhydrazone (FCCP), respectively) and between the tumour and normal sample. BAD, BIM, HRK and PUMA denote BH3-only BCL2 proteins, and the values (0.3, 1, 100) represent micromolar

concentrations. **c**, Example comparison of peptide-induced depolarization in a tumour and normal sample. Mean depolarization across three replicates for the tumour (purple) and normal (grey) samples are shown. The tumour shows significantly more mitochondrial priming than the normal sample ($P = 2.17 \times 10^{-5}$; paired *t*-test). **d**, BH3 peptide profiling in tumour–normal pairs. Shown is the mean depolarization across replicates for seven tumour and normal samples subjected to each peptide at the maximum concentration. Matched tumour (purple) and normal (grey) samples are connected in red. Error bars in **c** and **d** denote s.d. * $P < 0.05$; ** $P < 0.01$; NS, not significant ($P > 0.05$); paired *t*-test.

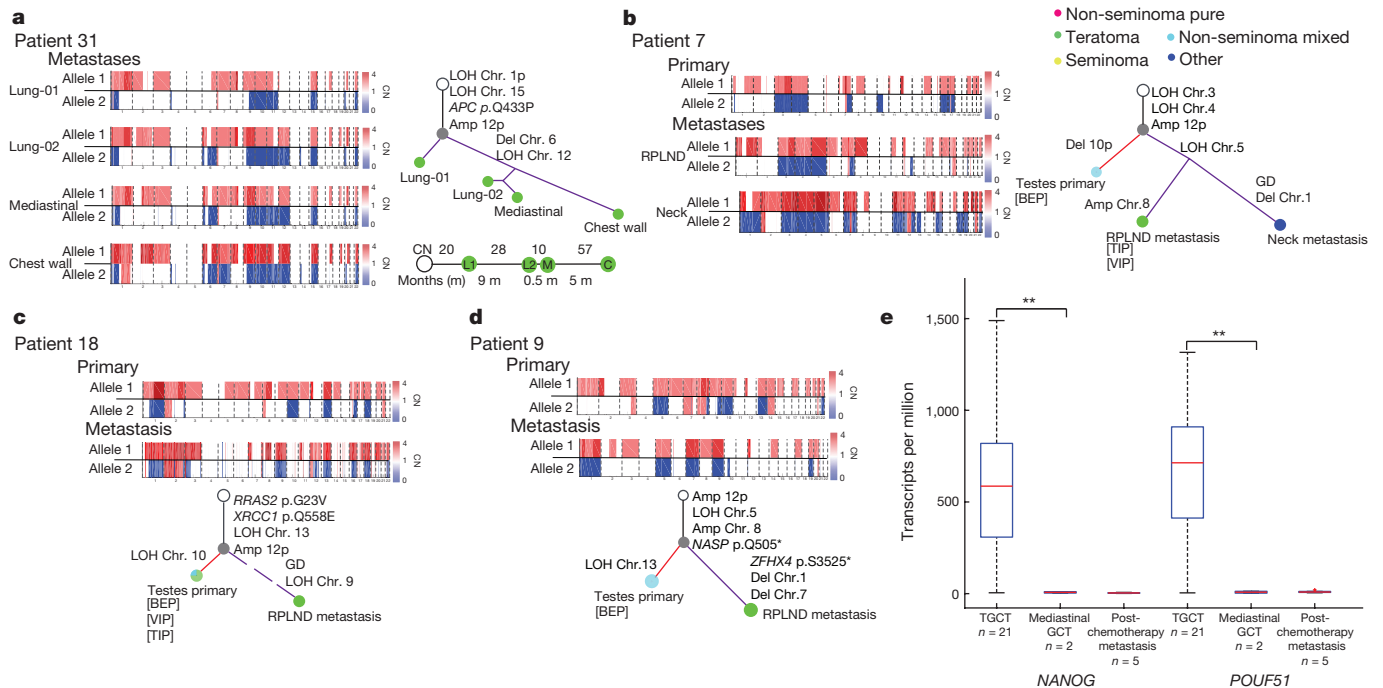


Figure 4 | Phylogenetic analysis and pluripotency of primary and metastatic GCTs. **a–d**, For each patient sample (DFCI_31 (**a**), DFCI_7 (**b**), DFCI_18 (**c**) and DFCI_9 (**d**)), histology proportion is indicated by pie charts within each phylogenetic tree. Treatments received after a given sample was acquired are listed below that respective sample (BEP, bleomycin, etoposide, cisplatin; TIP, paclitaxel, ifosfamide, cisplatin; VIP, vinblastine, ifosfamide, cisplatin). GD, genome doubling; RPLND, retroperitoneal lymph node dissection. Phylogenetic trees were constructed using allelic copy number deconstructions (Methods). Branches leading to primary samples are red,

and branches leading to metastases are purple. Dotted branches indicate deconstructions impacted by formalin-fixed, paraffin-embedded (FFPE) sample degradation, limiting length estimation. For patient DFCI_31 (**a**) the timeline below the tree shows time (months) and number of copy number events called by allelic deconstruction between samples (C, chest wall; L1, lung-01; L2, lung-02; M, mediastinal). **e**, Gene expression quantification of *NANOG* and *POU5F1* in primary and metastatic GCTs. Values are in transcripts per million. The red point denotes a *TP53*-mutant mediastinal primary GCT (DFCI_19). ** $P < 0.01$, Mann–Whitney test.

molecular characterization paired with mitochondrial priming studies of treatment-resistant tumours may inform features exclusive to this rare patient subset with significant unmet medical need for new therapies. This strategy may also inform strategies to improve the efficacy of chemotherapy in other malignancies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 May; accepted 2 November 2016.

- Oosterhuis, J. W. & Looijenga, L. H. Testicular germ-cell tumours in a broader perspective. *Nat. Rev. Cancer* **5**, 210–222 (2005).
- Sonne, S. B. *et al.* Analysis of gene expression profiles of microdissected cell populations indicates that testicular carcinoma in situ is an arrested gonocyte. *Cancer Res.* **69**, 5241–5250 (2009).
- Hanna, N. H. & Einhorn, L. H. Testicular cancer—discoveries and updates. *N. Engl. J. Med.* **371**, 2005–2016 (2014).
- Atkin, N. B. & Baker, M. C. Specific chromosome change, i(12p), in testicular tumours? *Lancet* **2**, 1349 (1982).
- Litchfield, K. *et al.* Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. *Nat. Commun.* **6**, 5973 (2015).
- Oosterhuis, J. W. *et al.* Ploidy of primary germ cell tumors of the testis. Pathogenetic and clinical relevance. *Lab. Invest.* **60**, 14–21 (1989).
- Montero, J. *et al.* Drug-induced death signaling strategy rapidly predicts cancer response to chemotherapy. *Cell* **160**, 977–989 (2015).
- Looijenga, L. H. *et al.* POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors. *Cancer Res.* **63**, 2244–2250 (2003).
- Hart, A. H. *et al.* The pluripotency homeobox gene *NANOG* is expressed in human germ cell tumors. *Cancer* **104**, 2092–2098 (2005).
- Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E. & Ulbright, T. M. The 2016 WHO classification of tumours of the urinary system and male genital organs—part A: renal, penile, and testicular tumours. *Eur. Urol.* **70**, 93–105 (2016).
- Gutekunst, M. *et al.* Cisplatin hypersensitivity of testicular germ cell tumors is determined by high constitutive Noxa levels mediated by Oct-4. *Cancer Res.* **73**, 1460–1469 (2013).
- Jacobsen, C. & Honecker, F. Cisplatin resistance in germ cell tumours: models and mechanisms. *Andrology* **3**, 111–121 (2015).
- Rijlaarsdam, M. A. & Looijenga, L. H. An oncofetal and developmental perspective on testicular germ cell cancer. *Semin. Cancer Biol.* **29**, 59–74 (2014).
- Rapley, E. A. *et al.* A genome-wide association study of testicular germ cell tumor. *Nat. Genet.* **41**, 807–810 (2009).
- Kanetsky, P. A. *et al.* Common variation in *KITLG* and at 5q31.3 predisposes to testicular germ cell cancer. *Nat. Genet.* **41**, 811–815 (2009).
- Feldman, D. R. *et al.* Presence of somatic mutations within PIK3CA, AKT, RAS, and FGFR3 but not BRAF in cisplatin-resistant germ cell tumors. *Clinical Cancer Res.* **20**, 3712–3720 (2014).
- Stachler, M. D. *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat. Genet.* **47**, 1047–1055 (2015).
- Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).
- Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
- Goudarzi, K. M. & Lindström, M. S. Role of ribosomal protein mutations in tumor development (Review). *Int. J. Oncol.* **48**, 1313–1324 (2016).
- Wang, L. *et al.* Novel somatic and germline mutations in intracranial germ cell tumours. *Nature* **511**, 241–245 (2014).
- Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
- Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
- Baker, D. E. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis *in vivo*. *Nat. Biotechnol.* **25**, 207–215 (2007).
- Närval, E. *et al.* High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. *Nat. Biotechnol.* **28**, 371–377 (2010).
- Liu, J. C. *et al.* High mitochondrial priming sensitizes hESCs to DNA-damage-induced apoptosis. *Cell Stem Cell* **13**, 483–491 (2013).
- Ni Chonghaile, T. *et al.* Pretreatment mitochondrial priming correlates with clinical response to cytotoxic chemotherapy. *Science* **334**, 1129–1133 (2011).
- Abada, P. B. & Howell, S. B. Cisplatin induces resistance by triggering differentiation of testicular embryonal carcinoma cells. *PLoS One* **9**, e87444 (2014).
- Wermann, H. *et al.* Global DNA methylation in fetal human germ cells and germ cell tumours: association with differentiation and cisplatin resistance. *J. Pathol.* **221**, 433–442 (2010).
- Jørgensen, A. *et al.* Dysregulation of the mitosis-meiosis switch in testicular carcinoma in situ. *J. Pathol.* **229**, 588–598 (2013).
- Hoffman, H. violin.m — Simple violin plot using Matlab default kernel estimation. INRES (Univ. of Bonn, 2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the patients for contributing to this study, and H.Taylor-Weiner for feedback on ES cells. This work was supported by NIH U54 HG003067, NIH 1K08 CA188615 (E.M.V.), Damon Runyon Clinical Investigator Award (E.M.V.), Shawmut Design and Construction Pan Mass Challenge Team (C.S.), and Giovino Jimmy Fund Golf Tournament (C.S.).

Author Contributions A.T.-W., T.Z., B.B., G.C.H., S.A., A.A.-M. and E.M.V. performed genomic analysis of discovery cohort. A.T.-W., T.Z., B.B., E.O., M.H., C.S. and E.M.V. performed clinical integration and analysis. J.L.G. and A.L. performed BH3 profiling experiments. S.S., S.L.C., R.B. and G.G. contributed methodology and analysis review. A.R. and E.M.V. performed biological review of genomic findings. S.G. performed sequencing assays. A.T.-W., T.Z., K.L., C.T. and E.M.V. performed genomic analysis of validation cohort. M.H. performed pathology and histological evaluation of clinical samples. A.T.-W., T.Z., B.B., C.S. and E.M.V. prepared manuscript and figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.M.V. (eliezerm_vanallen@dfci.harvard.edu).

Reviewer Information *Nature* thanks K. Nathanson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Patient enrollment. Patients with GCTs with available tumour and germline samples provided informed consent to an Institutional Review Board protocol that allows research molecular characterization of tumour and germline samples (Dana-Farber Cancer Institute 10-025 and 11-104). All available cases that met these criteria were included in this study. The clinical characteristics for these patients are available in Supplementary Tables 1 and 2.

Pathology review and sample processing. All samples were reviewed by a genitourinary oncology pathologist (M.H.). After fixation and mounting, 5–10 10- μ m slices from formalin-fixed, paraffin-embedded (FFPE) tumour blocks were obtained, and tumour-enriched tissue was macrodissected. Paraffin was removed from FFPE sections and cores using CitriSolv (Fisher Scientific), followed by ethanol washes and tissue lysis overnight at 56 °C. Samples were then incubated at 90 °C to remove DNA crosslinks, and DNA (and when possible, RNA) extraction was performed using Qiagen AllPrep DNA/RNA Mini Kit (51306). Germline DNA was obtained from adjacent normal tissue or peripheral blood mononuclear cells. **WES.** WES from FFPE tumour samples and matched germline DNA was performed as previously described³².

Library construction. DNA libraries for massively parallel sequencing were generated as previously described³² with the following modifications: the initial genomic DNA input into the shearing step was reduced from 3 μ g to 10–100 ng in 50 μ l of solution. For adaptor ligation, Illumina paired-end adapters were replaced with palindromic forked adapters (purchased from Integrated DNA Technologies) with unique 8-base index molecular barcode sequences included in the adaptor sequence to facilitate downstream pooling. With the exception of the palindromic forked adapters, all reagents used for end repair, A-base addition, adaptor ligation, and library enrichment PCR were purchased from KAPA Biosciences in 96-reaction kits. In addition, during the post-enrichment solid phase reversible immobilization bead cleanup, elution volume was reduced to 20 μ l to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted from the beads. Libraries with concentrations above 40 ng μ l⁻¹, as measured by a PicoGreen assay automated on an Agilent Bravo instrument, were considered acceptable for hybrid selection and sequencing.

Solution-phase hybrid selection. The exon capture procedure was performed as previously described³²; before hybridization, any libraries with concentrations >60 ng μ l⁻¹ (as determined by PicoGreen) were brought to 60 ng μ l⁻¹, and 8.3 μ l of library was combined with blocking agent, bait, and hybridization buffer. Libraries with concentrations between 50 and 60 ng μ l⁻¹ were normalized to 50 ng μ l⁻¹, and 10.3 μ l of library was combined with blocking agent, bait, and hybridization buffer. Libraries with concentrations between 40 and 50 ng μ l⁻¹ were normalized to 40 ng μ l⁻¹, and 12.3 μ l of library was combined with blocking agent, bait, and hybridization buffer. Finally, the hybridization reaction was reduced to 17 h; with no changes to the downstream capture protocol.

Preparation of libraries for cluster amplification and sequencing. After post-capture enrichment, libraries were quantified using PicoGreen, normalized to equal concentration using a Perkin Elmer MiniJanus instrument, and pooled by equal volume on the Agilent Bravo platform. Library pools were then quantified using quantitative PCR (KAPA Biosystems) with probes specific to the ends of the adapters; this assay was automated using Agilent's Bravo liquid handling platform. On the basis of quantitative PCR (qPCR) quantification, libraries were brought to 2 nM and denatured using 0.2 N NaOH on the Perkin-Elmer MiniJanus. After denaturation, libraries were diluted to 20 pM using hybridization buffer purchased from Illumina.

Cluster amplification and sequencing. Cluster amplification of denatured templates was performed according to the manufacturer's protocol (Illumina), HiSeq v3 cluster chemistry and flowcells, as well as Illumina's Multiplexing Sequencing Primer Kit, were used. DNAs were added to flowcells and sequenced using the HiSeq 2000 v3 Sequencing-by-Synthesis method, then analysed using RTA v1.12.4.2 or later. Each pool of whole exome libraries was subjected to paired 76 bp runs. An 8-base index-sequencing read was performed to read molecular indices, across the number of lanes needed to meet coverage for all libraries in the pool.

Alignment/assembly. Exome sequence data processing was performed using established analytical pipelines at the Broad Institute. A BAM file was produced with the Picard pipeline (<http://picard.sourceforge.net/>), which aligns the tumour and normal sequences to the hg19 human genome build using Illumina sequencing reads. The BAM was uploaded into the Firehose pipeline (<http://www.broadinstitute.org/cancer/cga/Firehose>), which manages input and output files to be executed.

Sequencing quality control. Quality control modules within Firehose were applied to all sequencing data for comparison of the origin for tumour and normal genotypes and to assess fingerprinting concordance. Cross-contamination of samples was estimated using ContEst³³. Samples with greater than 5% contamination or tumour purity less than 15% were excluded. All BAM files for this cohort are shared in dbGaP (phs000923.v1.p1). BAM files from the ref. 5 cohort were obtained through the European Genome-Phenome Archive (EGAS00001001084).

Mutation calling and significance analysis. MuTest³⁴ was applied to identify somatic single-nucleotide variants. Strelka³⁵ was applied to identify small insertions or deletions. Artefacts introduced by DNA oxidation during sequencing were computationally removed using a filter-based method³⁶. Mutations with allelic fractions of less than 0.05 excluded. Annotation of identified variants was done using Oncotator³⁷. Mutational significance analysis was performed using MutSigCV³⁸. Manual review of mutations in putative significantly mutated genes was subsequently performed using Integrated Genomics Viewer³⁹.

To compare mutations between distinct samples from the same patient, we used a previously described method¹⁷ designed to recover evidence for mutations called in one sample in all other samples derived from the same individual. In brief, the 'force-calling' method uses the strong prior of the mutation being present in at least one sample in the patient to more sensitively detect and recover mutations that might otherwise be missed.

Copy number and phylogenetic analysis. Copy-ratio profiles were inferred using ReCAPSEG (<http://gatkforums.broadinstitute.org/gatk/categories/recapseg-documentation>). Read depth at capture probes in tumour samples was normalized using a panel of normal samples to model noise and other biases. The resulting normalized copy ratios are then segmented using the circular binary segmentation algorithm⁴⁰. These data were then transformed into allelic copy number data via integration of allele fraction data from informative germline SNPs. Allelic copy number data allow for inference of contribution of each homologous allele to observed copy number shifts as well as identification of copy neutral events^{17,18}. Finally allelic copy number data was integrated with data from point mutations and short deletions and insertions as input to ABSOLUTE²².

The ABSOLUTE algorithm was used as previously described to generate purity and ploidy solutions for each tumour samples (Supplementary Table 6). These solutions provided estimates of total allelic copy number and cancer cell fraction for mutations and copy number events in the tumour samples. ABSOLUTE solutions were selected using manual curation. Using methods previously described²³, we performed allelic deconstruction using ABSOLUTE copy number data. In brief, deconstructions are performed using a Bayes approach in which likelihood of event progressions are defined probabilistically. We defined copy number events as arm level if the event spanned at least 80% of that arm and affected at least one allele. Events were described as reciprocal if homologous chromosomes were amplified and deleted. We used the allelic copy number deconstructions to draw phylogenetic trees. A copy number event was considered shared between two related samples if each of those samples contained an event whose start points fell within two probes of one another, whose end points fell within two probes of one another, and whose event amplitude was identical. Arm lengths are proportional to number of events delineating the samples on the ends of the branch.

We performed a permutation test of arm level deletions to determine the empirical significance of the observed rate of reciprocal deletions in our data set as previously described²². In brief, arm level deletions were shuffled between arms while keeping total deletions per sample constant. For each permutation the number of reciprocally deleted chromosome arms were counted. In 10,000 iterations we did not observe the number of reciprocal events in our data set.

Whole transcriptome sequencing and analysis: cDNA library construction. Total RNA was assessed for quality using the Caliper LabChip GX2. The percentage of fragments with a size greater than 200 nucleotides (DV200) was calculated using software. An aliquot of 200 ng of RNA was used as the input for first strand cDNA synthesis using Illumina's TruSeq RNA Access Library Prep Kit. Synthesis of the second strand of cDNA was followed by indexed adaptor ligation. Subsequent PCR amplification enriched for adapted fragments. The amplified libraries were quantified using an automated PicoGreen assay. Approximately 200 ng of each cDNA library, not including controls, was combined into 4-plex pools. Capture probes that target the exome were added, and hybridized for recommended time. After hybridization, streptavidin magnetic beads were used to capture the library-bound probes from the previous step. Two wash steps effectively remove any non-specifically bound products. These same hybridization, capture and wash steps are repeated to assure high specificity. A second round of amplification enriches the captured libraries. After enrichment the libraries were quantified with qPCR using the KAPA Library Quantification Kit for Illumina Sequencing Platforms and then pooled equimolarly. The entire process is in 96-well format and all pipetting is done by either Agilent Bravo or Hamilton Starlet.

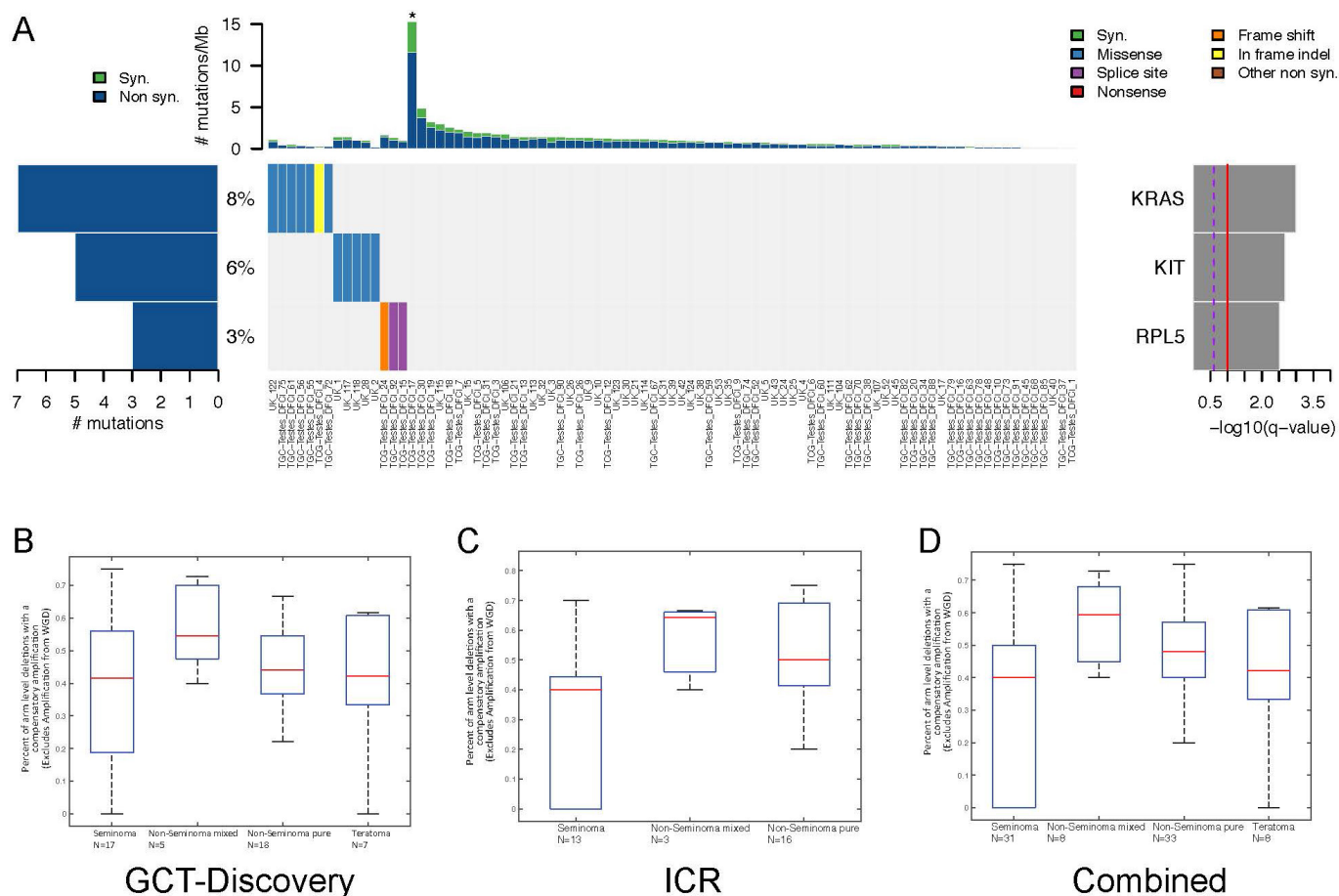
Illumina sequencing. Pooled libraries were normalized to 2 nM and denatured using 0.1 N NaOH before sequencing. Flowcell cluster amplification and sequencing were performed according to the manufacturer's protocols using HiSeq 2500. Each run was a 76-bp paired-end with an 8-base index barcode read. Data was analysed using the Broad Picard Pipeline which includes de-multiplexing and data aggregation.

Analysis. All BAMs were deduplicated using Samtools⁴¹ before gene expression quantification. Gene expression quantification was performed using RSEM⁴² to generate transcripts per million. Comparison of gene expression was performed for *POU5F1* and *NANOG* with Mann–Whitney test.

BH3 profiling. Normal and cancer testicular tissue was obtained from the Brigham and Women's Hospital Pathology Department (M.H.). Normal and tumour tissue was immediately processed. Tissue was chopped with a razor blade for 2 min until the tissue was a fine paste. It was then digested in trypsin with 10 mg ml⁻¹ collagenase IV and 10 U ml⁻¹ DNase I. The tissue was then incubated in digestion buffer at 37 °C for 21 min, vortexing hard every 7 min. The tissue was resuspended in full media (DMEM supplemented with 10% FBS, 2 mM L-glutamine, 100 U ml⁻¹ penicillin, and 100 µg ml⁻¹ streptomycin) and passed through a 70 µm filter to generate a single-cell suspension. The cells were spun down, washed once with PBS, and red blood cells were lysed using RBC lysis buffer. Cells were washed again with PBS, counted, and then used for BH3 profiling analysis^{27,43}.

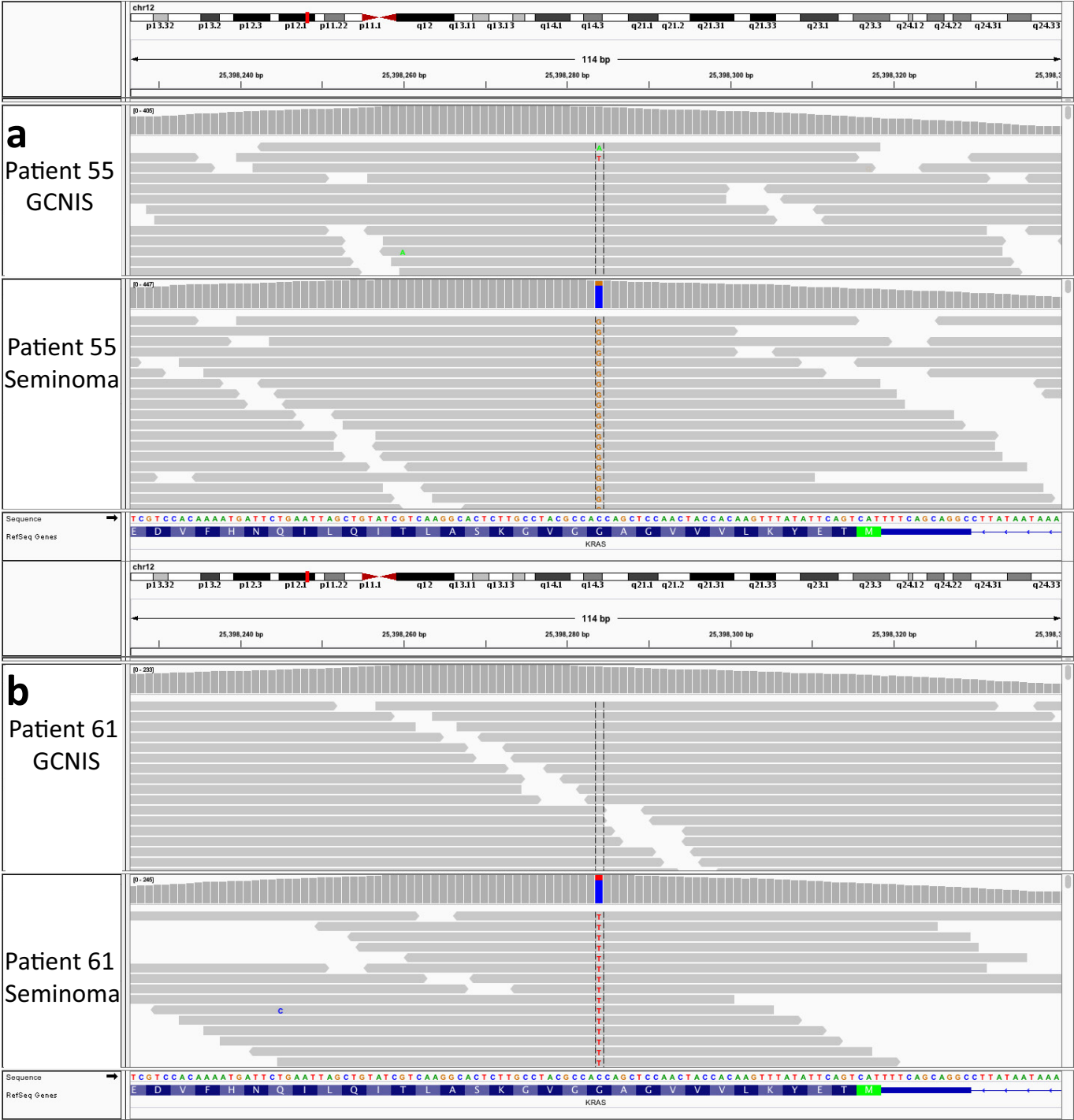
Data availability statement. BAM files are deposited in dbGaP (phs000923.v1.p1).

32. Van Allen, E. M. *et al.* Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat. Med.* **20**, 682–688 (2014).
33. Cibulskis, K. *et al.* ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* **27**, 2601–2602 (2011).
34. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
35. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
36. Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41**, e67 (2013).
37. Ramos, A. H. *et al.* Oncotator: cancer variant annotation tool. *Hum. Mutat.* **36**, E2423–E2429 (2015).
38. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
39. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
40. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
41. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
42. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
43. Touzeau, C. *et al.* BH3 profiling identifies heterogeneous dependency on Bcl-2 family members in multiple myeloma and predicts sensitivity to BH3 mimetics. *Leukemia* **30**, 761–764 (2016).



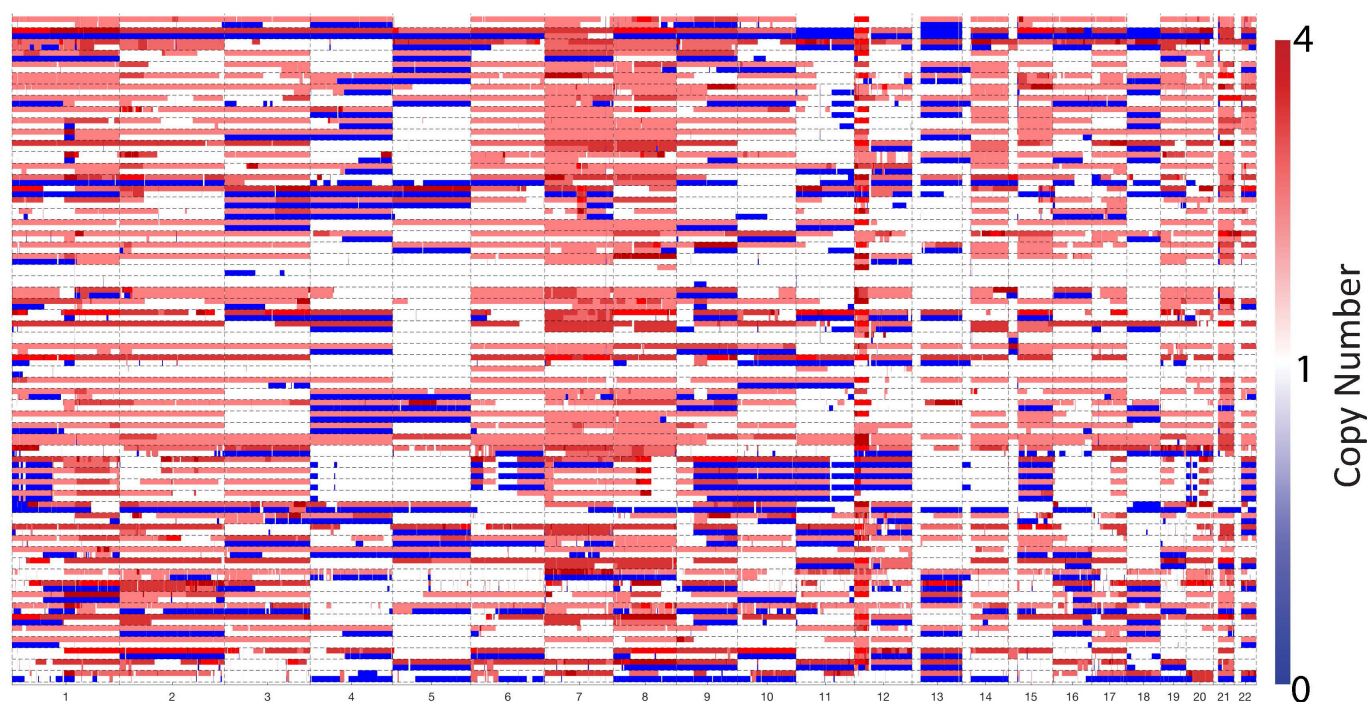
Extended Data Figure 1 | Mutational significance and copy number meta-analysis. a, Mutational significance meta-analysis of discovery and ICR cohorts identify *KRAS*, *KIT* and *RPL5* as significantly mutated in TGCT, with a spectrum of mutation rates. In this plot, each column

represents a patient WES. Asterisk denotes the hypermutated PMGCT (DFCI_17). **b**, RLOH distribution by histology in discovery cohort. **c**, RLOH distribution by histology in ICR cohort. **d**, RLOH distribution in the meta-analysis, consistent with both subsets.



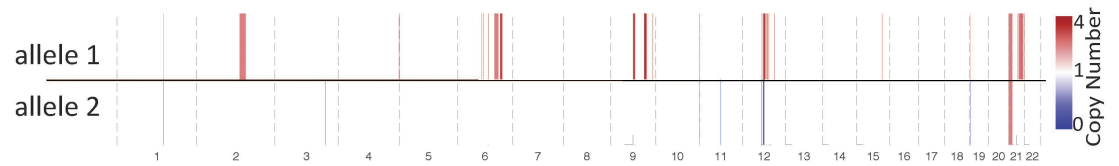
Extended Data Figure 2 | Genomic reads for *KRAS* loci in two patient cases. a, Integrative genomics viewer snapshot of *KRAS* p.G12A mutation in DFCI_55 GCNIS and seminoma. The mutation is present in the primary tumour but absent from the GCNIS. **b**, Integrative genomics

viewer snapshot of *KRAS* p.G12A mutation in DFCI_61 GCNIS and seminoma. The mutation is present in the primary tumour but absent from the GCNIS.

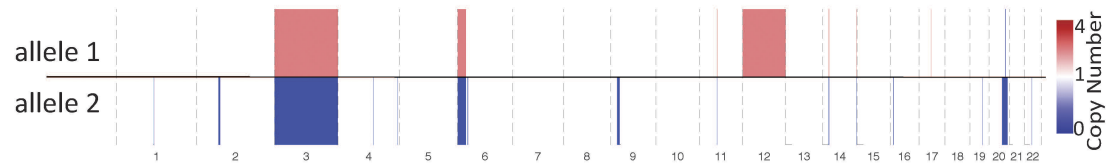


Extended Data Figure 3 | Allelic copy number heat map of the discovery cohort. Each tumour sample is a row, and chromosomes are listed as columns. Blue regions note deletions, and red regions denote amplifications.

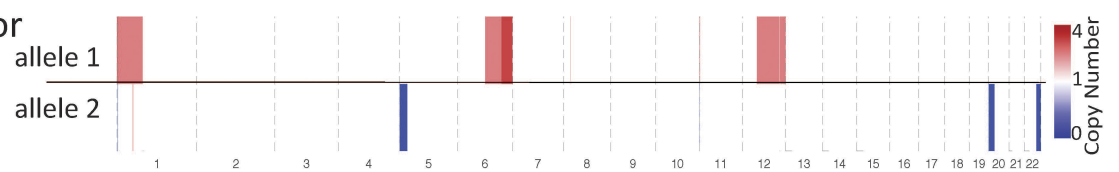
Liposarcoma DFCI 65



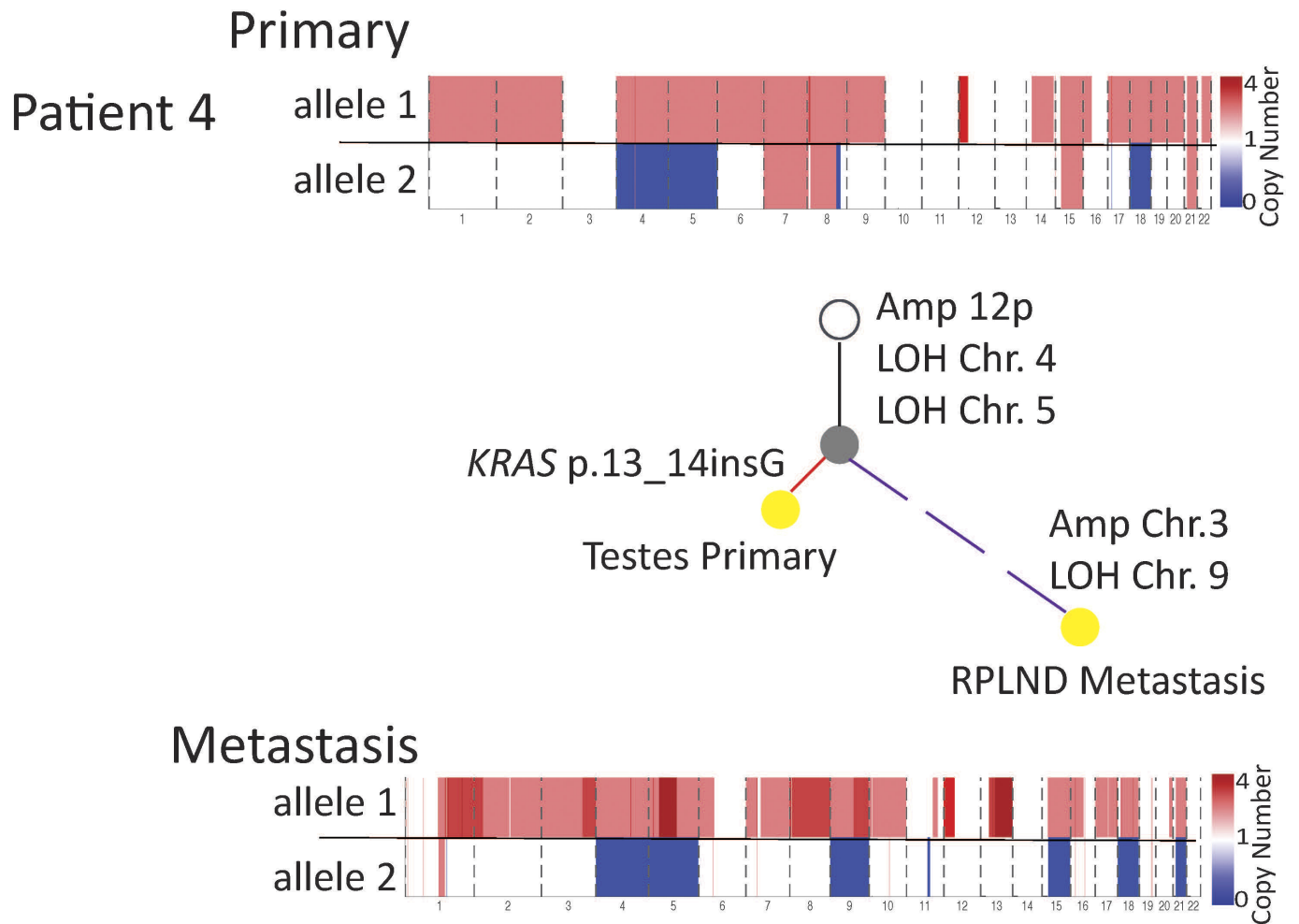
DLBCL DFCI 42



Sertoli cell tumor DFCI 39



Extended Data Figure 4 | Testes tumours of different cell types. Allelic copy number data from testes tumours of different cell types are shown. These three tumours do not contain the same level of arm level chromosomal events as GCTs.



Extended Data Figure 5 | Phylogenetic analysis of DFCI_4. Histology proportion is indicated by pie charts within each phylogenetic tree. Phylogenetic trees were constructed using allelic copy number deconstructions. Branch lengths are proportional to the number of

deconstructed copy number events. Branches leading to primary samples are red, and branches leading to metastases are purple. The dotted branch indicates deconstructions which may be impacted by FFPE sample degradation, limiting discrete branch length estimation.

Inhibition of mTOR induces a paused pluripotent state

Aydan Bulut-Karslioglu¹, Steffen Biechele¹, Hu Jin^{2,3}, Trisha A. Macrae¹, Miroslav Hejna^{2,3}, Marina Gertsenstein⁴, Jun S. Song^{2,3} & Miguel Ramalho-Santos¹

Cultured pluripotent stem cells are a cornerstone of regenerative medicine owing to their ability to give rise to all cell types of the body. Although pluripotent stem cells can be propagated indefinitely *in vitro*, pluripotency is paradoxically a transient state *in vivo*, lasting 2–3 days around the time of blastocyst implantation¹. The exception to this rule is embryonic diapause, a reversible state of suspended development triggered by unfavourable conditions². Diapause is a physiological reproductive strategy widely employed across the animal kingdom, including in mammals, but its regulation remains poorly understood. Here we report that the partial inhibition of mechanistic target of rapamycin (mTOR), a major nutrient sensor and promoter of growth³, induces reversible pausing of mouse blastocyst development and allows their prolonged culture *ex vivo*. Paused blastocysts remain pluripotent and competent—able to give rise to embryonic stem (ES) cells and live, fertile mice. We show that both naturally diapaused blastocysts *in vivo* and paused blastocysts *ex vivo* display pronounced reductions in mTOR activity, translation, histone modifications associated with gene activity and transcription. Pausing can be induced directly in cultured ES cells and sustained for weeks without appreciable cell death or deviations from cell cycle distributions. We show that paused ES cells display a remarkable global suppression of transcription, maintain a gene expression signature of diapaused blastocysts and remain pluripotent. These results uncover a new pluripotent stem cell state corresponding to the epiblast of the diapaused blastocyst and indicate that mTOR regulates developmental timing at the peri-implantation stage. Our findings have implications in the fields of assisted reproduction, regenerative medicine, cancer, metabolic disorders and ageing.

Preimplantation mammalian embryos can develop in suspension culture *ex vivo* up to the blastocyst stage. The E3.5 mouse blastocyst collapses after ~24–48 h (Fig. 1a), but its survival can be extended for several days if nutrients like glucose or certain amino acids are removed from the medium^{4,5}. We hypothesized that inhibiting growth pathways might induce a viable dormant state in blastocysts. We isolated blastocysts and cultured them *ex vivo* in the presence of small molecule inhibitors of translation, mTOR signalling, Myc family transcription factors or histone acetyltransferases (HATs) (Fig. 1a, b). We found that inhibition of translation, Myc or HATs has minimal effects on blastocyst survival, prolonging it by a maximum of 1 day relative to controls (Fig. 1b, Extended Data Fig. 1c–f). These results are in agreement with recent findings describing culture of Myc-depleted blastocysts for 18 h⁶. Remarkably, reducing mTOR activity using INK128 (ref. 7) enables a major extension of blastocyst culture by 9–12 days (equivalent days of gestation (EDG) 12.5–15.5) for the majority of embryos, reaching a maximum of 22 days *ex vivo*, that is, EDG25.5 (Fig. 1b). Given that mouse gestation lasts 19 days, these data indicate that blastocysts can be maintained in culture past the time it would take for birth to occur.

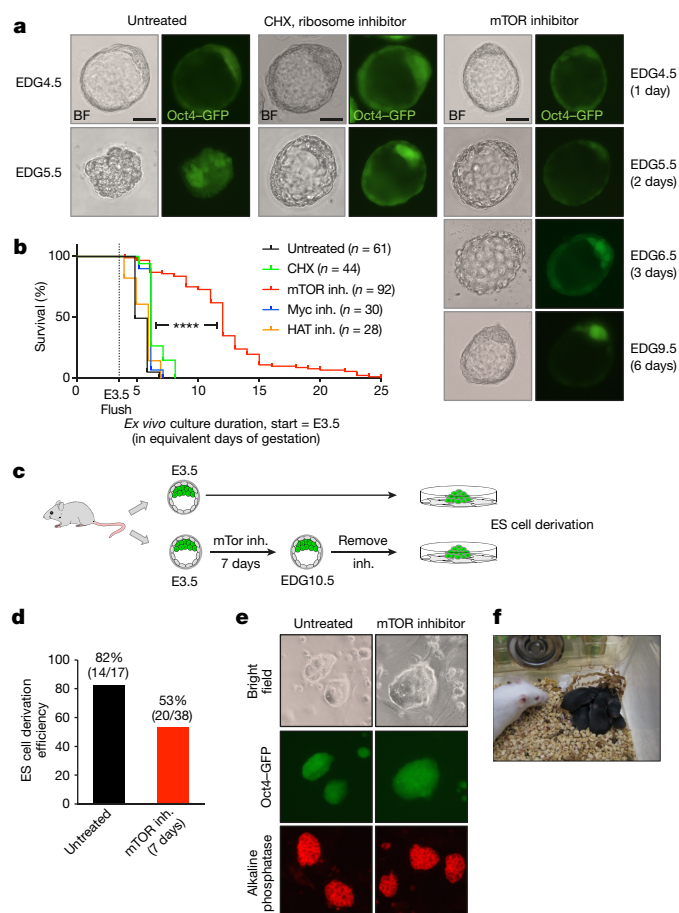


Figure 1 | mTOR inhibition induces blastocyst pausing *ex vivo*.

a, Representative images of bright field and Oct4-GFP expression of blastocysts cultured *ex vivo* under indicated conditions. Scale bars, 25 μ m. **b**, Kaplan-Meier survival curves of blastocysts cultured *ex vivo* with inhibitors. *n*, number of blastocysts tested. *****P* < 0.0001 from log-rank (Mantel-Cox) test. **c**, Workflow of ES cell derivation from untreated or mTOR-inhibited embryos. **d**, ES cell derivation efficiency from untreated and mTOR-inhibited embryos. The reduced efficiency of derivation from mTOR-inhibited embryos may be due to a lower frequency of hatching from the zona pellucida. **e**, Representative images of bright field, Oct4-GFP (passage 2) and alkaline phosphatase staining (passage 4) of ES cells derived from untreated and mTOR-inhibited embryos. Original magnification, $\times 200$. **f**, Live-born mice generated by transfer of mTOR-inhibited blastocysts at EDG8.5 into surrogate females. See Extended Data Fig. 2a for details.

¹Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, Center for Reproductive Sciences and Diabetes Center, University of California, San Francisco, San Francisco, California 94143, USA. ²Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana-Champaign, Urbana, Illinois 61801, USA. ³Departments of Bioengineering and Physics, University of Illinois, Urbana-Champaign, Urbana, Illinois 61801, USA. ⁴The Centre for Phenogenomics (TCP), Toronto, Ontario M5T 3H7, Canada.

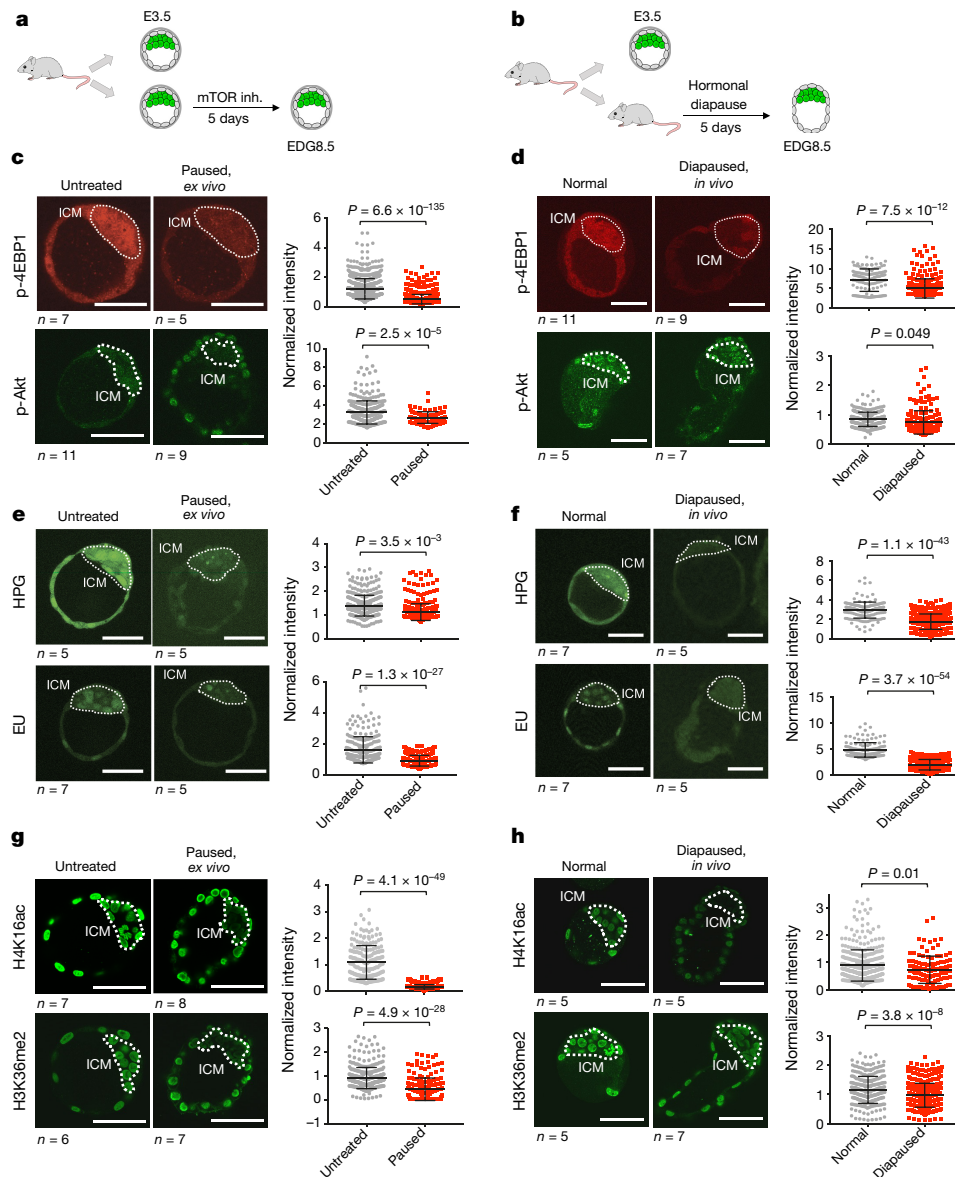


Figure 2 | *Ex vivo* paused blastocysts and *in vivo* diapaused blastocysts have similarly suppressed cellular functions. a, b, Workflow of the generation of *ex vivo* paused (a) and *in vivo* diapaused (b) blastocysts at EDG8.5. c–h, *Ex vivo* paused (c, e and g) and *in vivo* diapaused blastocysts (d, f and h) were stained for the indicated markers. c, d, 4EBP1 and Akt phosphorylation; e, f, nascent translation (HPG) or transcription (EU);

g, h, H4K16ac and H3K36me2. Representative images are shown. ICM is marked with dashed lines on the basis of Oct4 staining. Each ICM cell was quantified and pooled data from embryos were plotted in adjacent graphs. n = number of embryos. Values represent mean \pm s.d. P values are from two-tailed unpaired Student's t -test. Scale bars, 50 μ m.

Another recently developed inhibitor of mTOR (RapaLink-1), which, like INK128, inhibits both mTORC1 and mTORC2 complexes⁸, greatly extends blastocyst survival *ex vivo* (Extended Data Fig. 1a–c). Allosteric inhibitors like rapamycin, which target just the mTORC1 complex, only marginally extend blastocyst survival (Extended Data Fig. 1a–c), suggesting that inhibition of both complexes is required for developmental pausing. mTOR-inhibited blastocysts retain a well-expanded blastocoel, activity of the *Oct4-GFP* transgene and normal expression patterns of Nanog and Rex1 (Fig. 1a, Extended Data Fig. 2a, b). Apoptosis markers are largely absent in the inner cell mass (ICM) but can be detected in the trophectoderm of mTOR-inhibited blastocysts (Extended Data Fig. 2c, d).

Blastocysts cultured for 7 days in mTOR inhibitor give rise to ES cells (Fig. 1c, d) that express pluripotency markers (Fig. 1e). Moreover, blastocysts cultured in mTOR inhibitor for 4–5 days can give rise to live-born, fertile mice (Fig. 1f, Extended Data Fig. 3a, b). These results indicate that mTOR inhibition induces and sustains a reversible

paused pluripotent state (hereafter referred to as 'paused'). Notably, cleavage-stage embryos cannot be paused by inhibition of mTOR and instead develop with a slight delay to the blastocyst stage (Extended Data Fig. 3c).

Mouse blastocysts can undergo diapause *in utero* for up to 2 weeks if the pregnant female is lactating, a state that can be simulated hormonally⁹. In fact, the first ES cell lines were derived from diapaused blastocysts¹⁰. We compared EDG8.5 blastocysts generated either *ex vivo* through mTOR inhibition or *in vivo* by hormonally inducing diapause in pregnant females⁹ (Fig. 2a, b). A primary function of mTOR is to induce high translational output in growing cells by directly phosphorylating and inactivating 4EBP1, a repressor of translation³. The levels of phospho-4EBP1, a target of mTORC1, are significantly reduced in paused and diapaused blastocysts relative to control blastocysts (Fig. 2c, d). Phospho-Akt, an mTORC2 target and positive regulator of proliferation and metabolism³, is also reduced in paused blastocysts, and to a lesser extent in diapaused blastocysts (Fig. 2c, d). Moreover,

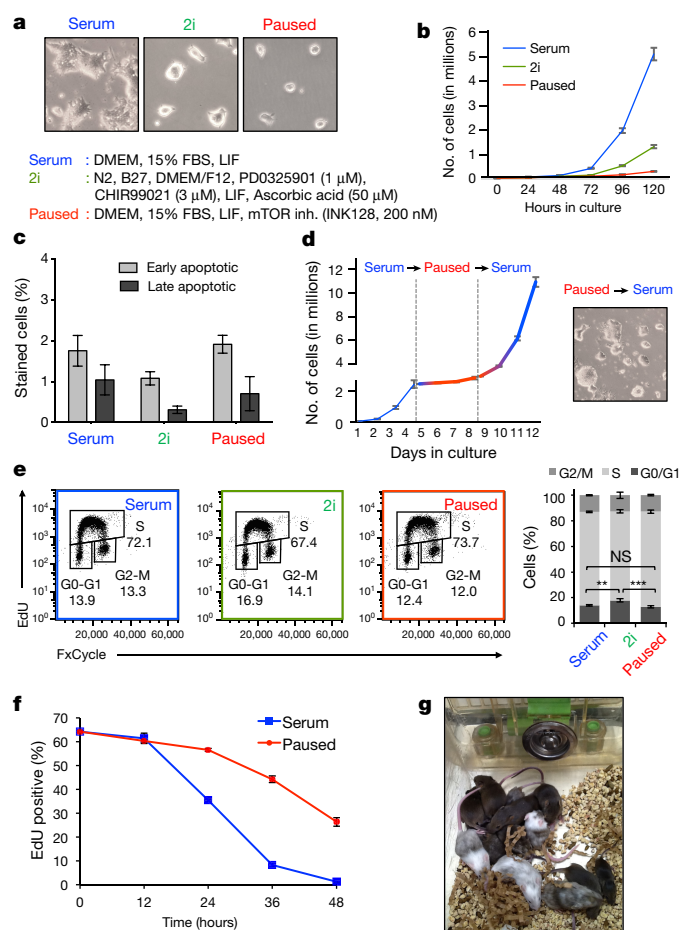


Figure 3 | mTOR inhibition induces pausing of ES cells. **a**, Representative bright field images of ES cells cultured in serum, 2i and paused conditions (original magnification, $\times 100$; media compositions listed). **b**, Expansion curves of serum, 2i and paused ES cells. Two biological replicates, each with three technical replicates, were performed. **c**, Flow cytometry analysis of apoptotic cell populations in serum, 2i and paused ES cells. Annexin V and Sytox Blue were used for early and late apoptosis. Eight technical replicates were performed. Error bars, s.d. **d**, Expansion curves of ES cells showing the reversible nature of the paused state. Two biological replicates, each with three technical replicates, were performed. **e**, Cell cycle analysis of serum, 2i and paused ES cells, based on four technical replicates of each state. Error bars, s.d. **f**, Flow cytometry showing dilution of incorporated EdU in serum and paused ES cells. Three technical replicates are shown. **g**, Chimeric mice generated from G4 ES cells previously paused for 7 days and released from pausing before morula aggregations with albino host embryos.

blastocysts paused *ex vivo* and diapaused *in vivo* display significant reductions in nascent protein synthesis relative to control embryos (Fig. 2e, f). These results are in agreement with previous reports showing reduced translation in diapaused blastocysts^{6,11}. However, suppressed translation alone is not sufficient to drive pausing, evidenced by the only slight extension in *ex vivo* blastocyst survival upon inhibition of protein synthesis (Fig. 1b).

mTOR is known to phosphorylate and inactivate regulators of autophagy. One major mTOR target in this context is Ulk1 (also known as Atg1)¹². Consistent with this notion, paused and diapaused blastocysts have reduced levels of phospho-Ulk1 (Extended Data Fig. 4a, b). Autophagy has previously been shown to be required for preimplantation development¹³ and implicated in blastocyst diapause *in vivo*¹⁴. We found that the survival ratio and longevity of paused blastocysts are significantly reduced when co-treated with the Ulk1 inhibitor SBI-0206965 (Extended Data Fig. 4c–e). Taken together with the published literature, these data suggest that autophagy is a component of blastocyst maintenance during developmental pausing.

We asked whether inhibition of growth and translation are associated with changes in transcriptional activity and chromatin landscape. We found that both paused and diapaused blastocysts display a marked reduction in nascent transcription (Fig. 2e, f). By immunofluorescence we did not detect significant changes in the global levels of H3K4me3, a chromatin mark associated with active promoters, or of H3K9me3, a heterochromatin mark (Extended Data Fig. 5b, c). By contrast, we found that both paused blastocysts and diapaused blastocysts exhibit sharply reduced levels of H4K16ac, H4K5/8/12ac and H3K36me2, particularly in the ICM (Fig. 2g, h, Extended Data Fig. 5a). Histone H4 acetylation and H3K36me2 are highly correlated with transcriptional activity. Thus, developmental pausing involves a chromatin landscape associated with suppressed transcription. Nevertheless, neither inhibition of histone acetylation nor suppression of Myc-mediated transcription^{15,16} can sustain a paused blastocyst state (Fig. 1b). Moreover, the trophectoderm and the ICM appear to differ in terms of changes to their chromatin landscapes in response to mTOR inhibition, yet both display suppressed transcription and translation in agreement with data implicating mTOR in trophectoderm differentiation¹⁷. These results suggest that a combination of multiple effects downstream of mTOR inhibition on both the ICM and the trophectoderm mediates blastocyst pausing.

Pluripotency exists in different states *in vivo*, which can be captured *in vitro* using distinct media formulations for ES cell culture¹. We sought to capture the paused pluripotent state *in vitro* by inhibiting mTOR. ES cells cultured in serum with mTOR inhibitor (paused) grow at a much slower rate than cells in either serum or 2i and vitamin C^{18,19} conditions (see Fig. 3b). Paused ES cells can be sustained in culture for weeks without appreciable cell death (Fig. 3c), unlike ES cells with reduced levels of Myc activity⁶, and rapidly resume growth upon release from the mTOR inhibitor (Fig. 3d).

Paused ES cells show the same cell cycle distribution as serum cells, whereas 2i cells have a slightly higher proportion of cells in G0/G1 (Fig. 3e). However, the 5-ethynyl-2'-deoxyuridine (EdU) label, which marks S phase, is diluted at a much slower rate in paused cells than in serum cells (Fig. 3f). These data suggest that paused ES cells progress more slowly through the cell cycle, without preferential accumulation at any particular stage. Similarly to paused and diapaused blastocysts, paused ES cells have reduced levels of mTORC1/mTORC2 activity and nascent translation, correlating with a decrease in cell size (Extended Data Fig. 6b–d). Upon release from pause, ES cells contribute to high-grade, germline-transmitting mouse chimaeras (Fig. 3g, Extended Data Fig. 6e), confirming their full developmental potential *in vivo*.

Paused ES cells display a global suppression of transcription, evident at both the total and nascent RNA levels (Fig. 4a, b). We next analysed the transcriptome of ES cells in serum, 2i and paused conditions. Owing to the significant differences in total RNA levels per cell between the three states, we first performed cell-number-normalized RNA sequencing (RNA-seq; see Methods, Extended Data Fig. 7, Supplementary Table 2). The analysis revealed a remarkably suppressed transcriptome in paused ES cells (Fig. 4c), in line with the hypotranscriptional state observed in paused and diapaused blastocysts. Notably, although the vast majority of genes are downregulated in the paused state relative to serum or 2i conditions, some transcriptional and translational repressors are selectively upregulated in paused ES cells (Fig. 4d). To further investigate the expression differences between the three states, we analysed gene expression levels relative to the total abundance of mRNA within each condition (that is, sequencing depth normalization, see Methods, Extended Data Fig. 8a). We found that the distribution of the transcriptome of paused ES cells is significantly biased towards a relative upregulation of repressors of transcription (Extended Data Fig. 8b–d, Supplementary Table 3). These results indicate that paused ES cells are characterized by a global state of hypotranscription, potentially mediated by a set of transcriptional repressors. These features make the paused ES cell transcriptome readily distinguishable from serum and 2i ES cells.

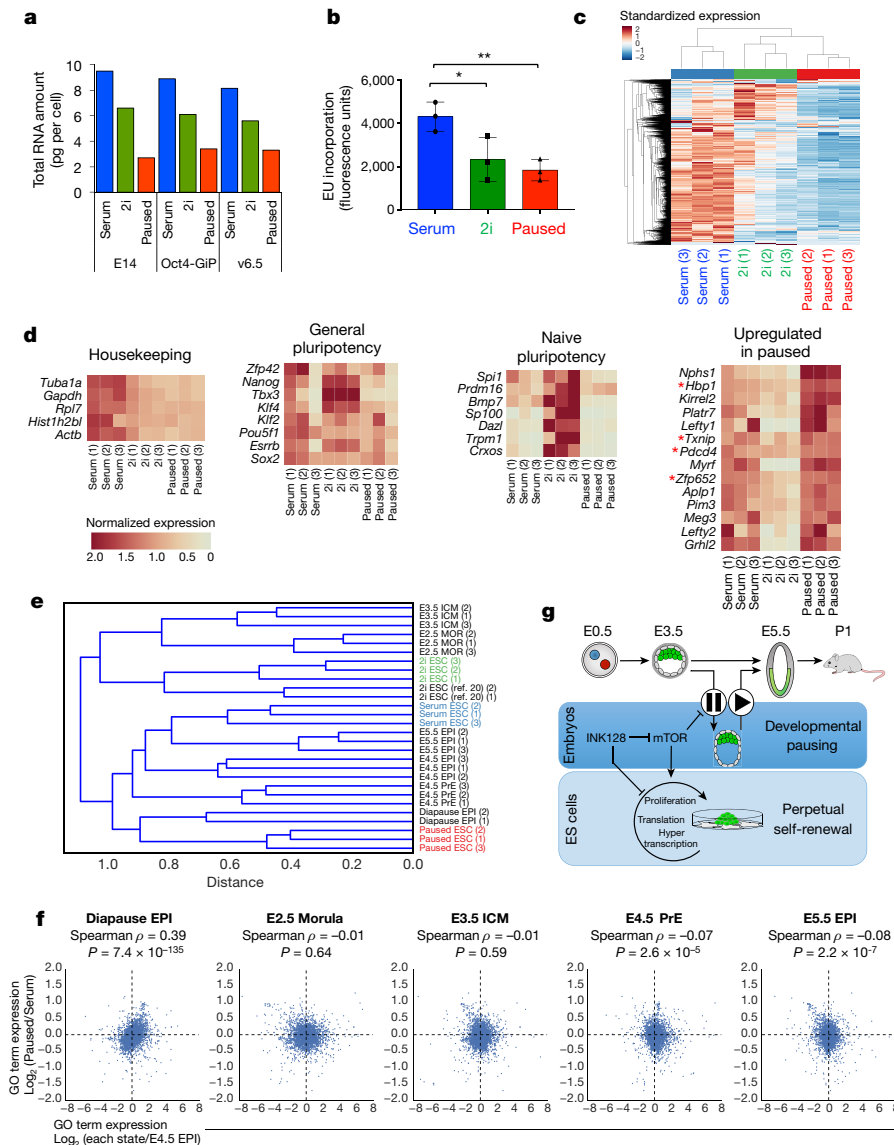


Figure 4 | Paused ES cells display global transcriptional suppression and mimic diapause. **a**, Total RNA amount per cell in three ES cell lines cultured in serum, 2i or pause conditions. The same cell lines were used for RNA-seq; numbers in parenthesis denote the biological replicates of each sample (**c–e**). **b**, Flow cytometry analysis of nascent transcription in the three states. Three technical replicates using E14 cells are shown. **c**, Hierarchical clustering showing standardized expression values for 5,992 genes with robust expression in at least one ES cell state (see Methods). **d**, Heat map representation of the indicated gene sets in serum, 2i and paused ES cells.

We next compared the transcriptomes of serum, 2i and paused ES cells to previously published gene expression data of various developmental stages, including diapause²⁰ (see Methods). This analysis revealed that paused ES cells have significant similarities with the diapaused epiblast but not with other developmental stages (Fig. 4e, f). These similarities are evident whether differential gene expression is analysed at the individual gene level or aggregated into gene ontology terms or annotated cellular pathways (Fig. 4f, Extended Data Fig. 9a, b). Unlike paused ES cells, the transcriptome of Myc-depleted cells⁶ is not consistently correlated with that of the diapaused epiblast (compare Fig. 4f and Extended Data Fig. 9 with Extended Data Fig. 10). Pathway analyses comparing paused ES cells, Myc-depleted ES cells and the diapause epiblast show that, while upregulated pathways are often concordant, several downregulated pathways are discordant in Myc-depleted ES cells (Extended Data Fig. 10e). Notably, pathways coordinately upregulated in paused ES cells and the diapaused epiblast are associated

Asterisks indicate repressor factors. **e**, **f**, Comparison of RNA-seq data from this study with distinct pluripotent cells *in vivo*²⁰. **e**, Hierarchical clustering indicating that the transcriptome of paused ES cells is similar to the *in vivo* diapaused epiblast. EPI, epiblast; MOR, morula. **f**, Scatter plots of gene ontology term expression. *y* axis represents log₂ fold change of gene ontology term expression in paused versus serum ES cells. *x* axes represent log₂ fold change of gene ontology term expression in the indicated developmental stages versus E4.5 epiblast. PrE, primitive endoderm. **g**, Model for the developmental pausing of blastocysts and ES cells by mTOR inhibition.

with sugar and lipid metabolism and are downregulated in diseases such as diabetes and immune disorders (Extended Data Fig. 9c). Taken together, our results document that paused blastocysts and paused ES cells induced by mTOR inhibition mimic natural diapause functionally and at the molecular level.

In summary, we show here that inhibition of mTOR allows the reversible suspension of developmental timing in mammalian embryos *ex vivo* (Fig. 4g). mTOR is essential for growth of mouse embryos at peri-implantation^{21–23}. Our results document that inhibiting mTOR captures a pluripotent stem cell state that, unlike previously described ES cell states, corresponds to a stage that can persist for weeks *in vivo*. These results pave the way for a genetic and molecular dissection of embryonic diapause. We speculate that the combinatorial effects of mTOR inhibition on transcription, translation and metabolism may be important for achieving developmental pausing. The ability to inhibit mTOR with small molecules allows direct comparisons

to other organisms and will contribute to studies of the evolution of developmental pausing²⁴. It will be of interest to explore the utility of manipulations of the mTOR pathway in assisted reproduction, regenerative medicine, preservation of cell viability after trauma and ageing. Moreover, the finding that mTOR inhibition induces a reversibly paused state in ES cells should be taken into consideration in cancer therapies that make use of mTOR inhibitors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 May; accepted 1 November 2016.

Published online 23 November 2016.

- Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).
- Fenelon, J. C., Banerjee, A. & Murphy, B. D. Embryonic diapause: development on hold. *Int. J. Dev. Biol.* **58**, 163–174 (2014).
- Zoncu, R., Efeyan, A. & Sabatini, D. M. mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat. Rev. Mol. Cell Biol.* **12**, 21–35 (2011).
- Naeslund, G. The effect of glucose-, arginine- and leucine-deprivation on mouse blastocyst outgrowth *in vitro*. *Ups. J. Med. Sci.* **84**, 9–20 (1979).
- Gwatkin, R. B. L. Amino acid requirements for attachment and outgrowth of the mouse blastocyst *in vitro*. *J. Cell. Physiol.* **68**, 335–343 (1966).
- Scognamiglio, R. *et al.* Myc depletion induces a pluripotent dormant state mimicking diapause. *Cell* **164**, 668–680 (2016).
- Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55–61 (2012).
- Rodrik-Outmezguine, V. S. *et al.* Overcoming mTOR resistance mutations with a new-generation mTOR inhibitor. *Nature* **534**, 272–276 (2016).
- MacLean Hunter, S. & Evans, M. Non-surgical method for the induction of delayed implantation and recovery of viable blastocysts in rats and mice by the use of tamoxifen and Depo-Provera. *Mol. Reprod. Dev.* **52**, 29–32 (1999).
- Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
- Fu, Z. *et al.* Integral proteomic analysis of blastocysts reveals key molecular machinery governing embryonic diapause and reactivation for implantation in mice. *Biol. Reprod.* **90**, 52–52 (2014).
- Chan, E. Y. mTORC1 phosphorylates the ULK1-mAtg13-FIP200 autophagy regulatory complex. *Sci. Signal.* **2**, pe51–pe51 (2009).
- Tsukamoto, S. *et al.* Autophagy is essential for preimplantation development of mouse embryos. *Science* **321**, 117–120 (2008).
- Lee, J.-E. *et al.* Autophagy regulates embryonic survival during delayed implantation. *Endocrinology* **152**, 2067–2075 (2011).
- Nie, Z. *et al.* c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* **151**, 68–79 (2012).
- Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).
- Martin, P. M. & Sutherland, A. E. Exogenous amino acids regulate trophectoderm differentiation in the mouse blastocyst through an mTOR-dependent pathway. *Dev. Biol.* **240**, 182–193 (2001).
- Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).
- Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
- Boroviak, T. *et al.* Lineage-specific profiling delineates the emergence and progression of naive pluripotency in mammalian embryogenesis. *Dev. Cell* **35**, 366–382 (2015).
- Murakami, M. *et al.* mTOR is essential for growth and proliferation in early mouse embryos and embryonic stem cells. *Mol. Cell. Biol.* **24**, 6710–6718 (2004).
- Guertin, D. A. *et al.* Ablation in mice of the mTORC components raptor, rictor, or mLST8 reveals that mTORC2 is required for signaling to Akt-FOXO and PKC α , but not S6K1. *Dev. Cell* **11**, 859–871 (2006).
- Gangloff, Y.-G. *et al.* Disruption of the mouse mTOR gene leads to early postimplantation lethality and prohibits embryonic stem cell development. *Mol. Cell. Biol.* **24**, 9508–9516 (2004).
- Schiesari, L. & O'Connor, M. B. Diapause: delaying the developmental clock in response to a changing environment. *Curr. Top. Dev. Biol.* **105**, 213–246 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to K. Shokat for the kind gift of Rapalink-1. We thank R. Blueloch, B. Bruneau, M. Conti, S. Fisher, D. Ruggero, and members of the Santos Laboratory for critical reading of the manuscript. This research was supported by grants NIH 5P30CA082103 to the UCSF Center for Advanced Technology, NIH P30DK063720 to the UCSF Flow Cytometry Core, NSF 1442504 and NIH R01CA163336 to J.S.S., and NIH R01OD012204 and R01GM113014 to M.R.-S.

Author Contributions A.B.-K., S.B. and M.R.-S. conceived of the project. A.B.-K. and S.B. isolated embryos. A.B.-K. performed most paused embryo and ES cell cultures, embryo stainings and quantifications, and ES cell experiments with the following exceptions: T.A.M. performed HPG, EU and cell cycle analyses on ES cells. S.B. generated diapaused embryos, performed all NSET embryo transfers, and parturitions when necessary. H.J., and M.H. analysed RNA-seq data under the supervision of J.S.S. M.G. performed parallel embryo culture for surgical transfers and ES cell culture for aggregations, supported by staff of TCP Model Production Core. M.R.-S. supervised the project. A.B.-K. and M.R.-S. wrote the manuscript with feedback from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R.-S. (mrsantos@ucsf.edu).

Reviewer Information *Nature* thanks K. Kaji, T. Rodríguez and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Animal studies. The following strains of mice were used (see details in following sections): Swiss Webster females and males, C57BL/6J or C57BL6/N males, B6.Cg-Tg(Pou5f1-GFP)1Scho²⁵ males, CD-1 females and males. 6–10-week-old female mice, and 6-week- to 6-month-old male mice were used. Animals were maintained on 12 h light–dark cycle and provided with food and water *ad libitum* in individually ventilated units (Techniplast at TCP, Laboratory Products at UCSF) in the specific-pathogen-free facilities at UCSF and at TCP. All procedures involving animals were performed in compliance with the protocol approved by the IACUC at UCSF, as part of an AAALAC-accredited care and use program (protocol AN091331-03); and according to the Animals for Research Act of Ontario and the Guidelines of the Canadian Council on Animal Care. Animal Care Committee reviewed and approved all procedures conducted on animals at TCP. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. No statistical methods were used to predetermine sample size estimate.

Embryo culture. Unless otherwise indicated, Swiss Webster females were mated to Swiss Webster males, or to C57BL/6 males homozygous for an *Oct4-GFP* transgene (B6.Cg-Tg(Pou5f1-GFP)1Scho)²⁵. Preimplantation embryos were collected at indicated time-points after detection of the copulatory plug by flushing oviducts (E1.5–E2.5) or uteri (E3.5) of pregnant females using M2 medium (Zenith Biotech) supplemented with 2% BSA (Sigma). Subsequent embryo culture was performed in 4-well plates in 5% O₂, 5% CO₂ at 37 °C in KSOM^{AA} Evolve medium (Zenith Biotech) with 2% BSA and the following inhibitors, after optimization of concentrations: 200 nM INK128 (Medchem Express), 2.5 µM 10058-F4 (Sigma), 100 ng ml^{−1} cycloheximide (Amresco), 50 µM Anacardic Acid (Sigma). Other mTOR inhibitors (AZD2014, Everolimus and Rapamycin (Medchem Express) and Rapalink-1 (gift of K. Shokat)) and autophagy inhibitors chloroquine (Sigma) and SBI-0206965 (Medchem Express) were used at the indicated concentrations under same culture conditions.

Diapause induction. Diapause was induced as previously described⁹ after natural mating of Swiss Webster mice. Briefly, pregnant females were injected at E2.5 and EDG5.5 with 10 µg tamoxifen (intra-peritoneally) and at E2.5 only with 3 mg medroxyprogesterone 17-acetate (subcutaneously). Diapaused blastocysts were flushed from uteri in M2 media after 4 days of diapause at EDG8.5.

Embryo transfer. Both surgical and non-surgical embryo transfers (NSET) were performed. For surgical transfers, superovulated CD-1 females were mated to C57BL/6J or C57BL6/N males and embryos were flushed at E3.5. Embryo culture (as described above) and surgical embryo transfer into the uteri of 2.5 days post coitus pseudopregnant CD-1 females previously mated with vasectomized CD-1 males was performed essentially as described²⁶. For NSET, Swiss Webster females were mated to vasectomized CD-1 males and transfer was performed at E2.5 of surrogate according to manufacturer's instructions (ParaTechs, Lexington). Before embryo transfer, embryos were cultured in KSOM^{AA}, 2% BSA without inhibitor for 1 h. In the cases indicated (Extended Data Fig. 1a), Caesarian delivery was performed at E20, followed by fostering to Swiss Webster females. Coat colour markers (agouti versus albino) were used to distinguish transferred embryos after birth.

ES cell derivation. ES cell derivation was performed as previously described²⁷. Swiss Webster females were naturally mated to Swiss Webster-C57BL/6 males heterozygous for an *Oct4-GFP* transgene (B6.Cg-Tg(Pou5f1-GFP)1Scho)²⁵. Blastocysts were collected by flushing uteri of pregnant females at E3.5, and were seeded on feeders either immediately or after culturing for 7 days in KSOM^{AA}, 2% BSA, 200 nM INK128. Imaging of fluorescence driven by the *Oct4-GFP* transgene and alkaline phosphatase activity (VECTOR Red AP Substrate Kit, Vector Laboratories) was performed using a Leica DM IRB microscope.

Embryo immunofluorescence. For immunofluorescence stainings, normal (E3.5), *in vivo* diapaused or *ex vivo* paused embryos were fixed in 4% paraformaldehyde for 15 min, washed with PBS and permeabilized with 0.2% Triton X-100 in PBS for 15 min. After blocking in PBS, 2.5% BSA, 5% donkey serum for 1 h, embryos were incubated overnight at 4 °C with the following primary antibodies in blocking solution: phospho-4EBP1 (Thr37/46, clone 236B4), phospho-Akt (Ser473), phospho-Ulk1 (Ser757), Nanog, c-Parp, c-Caspase3 (all from Cell Signaling), H3K4me3, H4K16ac, H4K5/8/12ac, H3K9me3 (all from Millipore), Oct4 and Rex1 (Santa Cruz Biotechnology) and H3K36me2 (Abcam). Embryos were washed in PBS-Tween20, 2.5% BSA, incubated with fluorescence-conjugated secondary antibodies (Invitrogen) for 2 h at room temperature, and mounted in VectaShield mounting medium with DAPI (Vector Laboratories). For labelling nascent transcription or translation, embryos were labelled in their respective culture medium for 20 min with EU (5-ethynyl uridine) or HPG (L-homopropargylglycine) following the manufacturer's instructions for Click-iT RNA and protein labelling kits (Thermo Fisher Scientific). Imaging was performed using a Leica SP5 confocal microscope with automated z-stacking at 10 µm intervals. Cell Profiler Software²⁸ was used for image quantification and Prism (Graphpad Software) was used for

plotting data points. Datasets do not show similar variance between control and paused/diapaused embryos in all cases, therefore we applied Welch's correction to the statistical analysis.

ES cell culture. E14 (from B. Skarnes, Sanger Institute), Oct4-GiP (from A. Smith, University of Cambridge) and v6.5 (from R. Blelloch, UCSF) ES cell lines were used. 'Serum' cells were cultured in ES-FBS medium: DMEM GlutaMAX with Na Pyruvate (Thermo Fisher Scientific), 15% FBS (Atlanta Biologicals), 0.1 mM non-essential amino acids, 50 U ml^{−1} penicillin/streptomycin (UCSF Cell Culture Facility), 0.1 mM EmbryoMax 2-Mercaptoethanol (Millipore) and 2,000 U ml^{−1} ESGRO supplement (LIF, Millipore). '2i' cells were cultured in ES-2i medium: DMEM/F-12, Neurobasal medium, 1 × N2/B27 supplements (Thermo Fisher Scientific), 1 µM PD0325901, 3 µM CHIR99021 (Selleck Chemicals), 50 µM Ascorbic acid (Sigma) and 2,000 U ml^{−1} ESGRO supplement (LIF) (Millipore). 'Paused' cells were cultured in ES-FBS medium containing 200 nM INK128 (Medchem). ES cells can also be paused in 2i medium, but the mTOR inhibitor needs to be removed at each passaging and reintroduced after colony formation to avoid major cell death (Extended Data Fig. 6a). The cell lines have not been authenticated. E14 and v6.5 tested negative for mycoplasma contamination. Oct4-GiP was not tested.

Generation of chimaeras. R1 (129S1 × 129X1)²⁹ and G4 (129S6 × B6N)³⁰ ES cells were used for morula aggregations. ES cells were cultured in DMEM containing 10% FBS (Wisent, lot-tested to support generation of germline chimaeras), 10% KnockOut Serum Replacement, 2 mM GlutaMAX, 1 mM Na Pyruvate, 0.1 mM non-essential amino acids, 0.1 mM 2-Mercaptoethanol (all Thermo Fisher Scientific), 1,000 U ml^{−1} LIF (Millipore). G4 ES cells were grown on MEF obtained from TgN(DR4)1Jae/J mice at all times except one passage on gelatinized tissue culture plates before aggregation. R1 ES cells were cultured in feeder-free conditions on gelatinized tissue culture plates. CD-1 (ICR) (Charles River) outbred albino stock was used as embryo donors for aggregation with ES cells and as pseudopregnant recipients. Details of morula aggregation can be found in²⁶. Briefly, embryos were collected at E2.5 from superovulated CD-1(ICR) female mice. Zona pellucidae of embryos were removed by the treatment with acid Tyrode's solution (Sigma). ES cell colonies were treated with 0.05% Trypsin-EDTA to lift loosely connected clumps. Each zona-free embryo was aggregated with 10–15 ES cells inside depression well made in the plastic dish with an aggregation needle (BLS Ltd, Hungary) and cultured overnight in microdrops of KSOM^{AA} covered by embryo-tested mineral oil (Zenith Biotech) at 37 °C in 94% air/6% CO₂. The following morning morulae and blastocysts were transferred into the uteri of E2.5 pseudopregnant CD-1(ICR) females previously mated with vasectomized males. Chimaeras were identified at birth by the presence of black eyes and later by the coat pigmentation. Chimeric males with more than 50% coat colour contribution were individually bred with CD-1(ICR) females. Germline transmission of ES cell genome was determined by eye pigmentation of pups at birth and later by the coat pigmentation.

Western blot analysis. 1 × 10⁶ cells were collected and lysed in RIPA buffer containing 1 × Protease Inhibitor Cocktail, 1 mM PMSF, 5 mM NaVO₄ and 5 mM NaF. Extracts were loaded into 4–15% Mini-Protein TGX SDS Page gels (Bio-Rad). Proteins were transferred to PVDF membranes. Membranes were blocked in 5% milk/PBS-T buffer for 30 min and incubated either overnight at 4 °C or 1 h at room temperature with the following antibodies: 4EBP1 (total or pThr37/46), S6K1 (total or pThr389), Akt (total or pSer473), mTOR (total or pSer2448) (Cell Signaling Technology), Gapdh (Millipore) and anti-rabbit/mouse secondary antibodies (Jackson Labs). Membranes were incubated with ECL or ECL Plus reagents and exposed to X-ray films (Thermo Fisher Scientific).

Cell cycle analysis. 4 × 10⁵ cells were seeded on 6-well plates. After overnight culture, cells were incubated for 1 h with 5-ethynyl-2-deoxyuridine (EdU) diluted to 10 µM in the indicated ES cell media. All samples were processed according to the manufacturer's instructions (Click-iT EdU Alexa Fluor 488 Imaging kit, Thermo Fisher Scientific). EdU incorporation was detected by Click-iT chemistry with an azide-modified Alexa Fluor 488. Cells were resuspended in EdU permeabilization/wash reagent and incubated for 30 min with FxCycle Violet Stain (Thermo Fisher Scientific). For EdU dilution experiments, ES cells were labelled for 90 min in serum, and afterwards were split into either serum or pause conditions; EdU analysis was done every 12 h for 48 h. Flow cytometric was performed on a LSRII flow cytometer (BD) and analysed using FlowJo v10.0.8. Data sets show similar variance.

Analysis of nascent transcription or translation. Total nascent transcription (Ethynyl Uridine, EU) or translation (L-homopropargylglycine, HPG) were assessed in ES cells using the Click-iT RNA Alexa Fluor 488 HCS Assay kit according to the manufacturer's instructions (Thermo Fisher Scientific). Samples were analysed on a BD LSRII. Datasets show similar variance.

Analysis of apoptosis. After overnight culture on a 96-well plate, ESCs were washed once with PBS and trypsinized to single cells. They were resuspended in 10 µl of Annexin V diluted 1:100 in Binding Buffer (BioLegend) and incubated for

10 min in the dark. Cells were resuspended in 90 μ l of binding buffer with Sytox Blue (Thermo Fisher Scientific) at 1:10,000. Data were collected on a BD LSR II. Datasets show similar variance.

Generation and sequencing of RNA-seq samples. Three replicates were used for all samples. Freshly collected single-cell suspensions were sorted on a FACSAria II cell sorter to collect 10^5 cells for each sample. Total RNA was isolated using the RNeasy kit (Qiagen). All samples were spiked-in with ERCC control RNAs (Thermo Fisher Scientific) following manufacturer's recommendations. mRNA isolation and library preparation were performed on 250 ng total RNA from all samples using NEBNext Ultra Directional RNA library prep kit for Illumina (New England Biolabs). Samples were sequenced at The Center for Advanced Technology, UCSF on Illumina HiSeq2500.

Mapping reads. Single-end 50-bp reads were mapped to the mm10 mouse reference genome using Tophat2 (ref. 31) with default parameters. We used Cuffnorm and Cuffdiff with the gtf file from UCSC mm10 (Illumina iGenomes July 17, 2015 version) as transcript annotation to evaluate relative expression level of genes (fragments per kilobase of transcript per million mapped reads (FPKM)) and call differentially expressed genes. The alignment rate exceeded 96% in all of our samples, yielding ~40 million aligned reads per sample. Data from ref. 20 and ref. 6 were downloaded from GEO and ArrayExpress, respectively, and processed with the same pipeline as our data.

Normalizing absolute expression. The absolute abundance of mRNA transcripts was estimated using the ERCC92 RNA spike-in³². ERCC92 contains 92 synthetic sequences with lengths ranging from 250 to 2,000 bp and concentration ranging over several orders of magnitude. ERCC sequences were designed to mimic mammalian mRNA, but are not homologous to the mouse genome, ensuring their unique mappability. We aligned the reads to the 92 reference spike-in sequences and compared the abundance of these sequences between different samples. As ERCC sequence abundances followed a highly linear trend in all pairs of samples across at least 5 orders of magnitude (Pearson correlation coefficient larger than 99.7%, see Extended Data Fig. 7), we assessed the absolute abundance of mRNA as the number of mRNA fragments per kilobase of transcript per 10 thousand mapped reads of ERCC. The overall abundance of ERCC spike-in sequences in our samples varied from 0.3% to 0.5% of aligned reads.

Suppressing batch effects and clustering. To facilitate better comparison between our data and data from ref. 20 and to reduce possible batch effects, in Fig. 4e, we followed the 'batch mean-centering' approach widely used in microarray gene expression data analysis for batch effect removal³³. Specifically, we separately mean-centred the $\log_2(\text{FPKM} + 1)$ value of each gene by subtracting the mean $\log_2(\text{FPKM} + 1)$ across all our samples (serum, 2i and paused) and across the samples from ref. 20. The numerical values of the mean-centred expression may not be directly comparable across all samples, because they may still have different dynamic ranges in different batches. We therefore used $1 - \text{Spearman correlation coefficient}$ as distance in the hierarchical clustering.

Clustering analysis. In Fig. 4c, we identified 5,992 genes with robust expression (cell-number-normalized expression value >50 in serum, 2i, or paused states). The cell-number-normalized expression value of each gene was standardized across the 9 samples by subtracting the mean and then dividing by the standard deviation. Hierarchical clustering was performed using the standardized expression values using Euclidean metric and average linkage. In Fig. 4e, in order to compare our samples with those from ref. 20, we used the $\log_2(\text{FPKM} + 1)$ value of each gene. Hierarchical clustering was performed using mean-centred (within each batch) expression values of 9,418 genes robustly expressed ($\text{FPKM} > 10$) in at least one cell state (serum, 2i, paused, diapause EPI, E2.5 MOR, E3.5 ICM, E4.5 EPI, E4.5

PrE, E5.5 EPI, or ESC 2i/LIF). $1 - \text{Spearman correlation coefficient}$ was used as distance and average linkage was used.

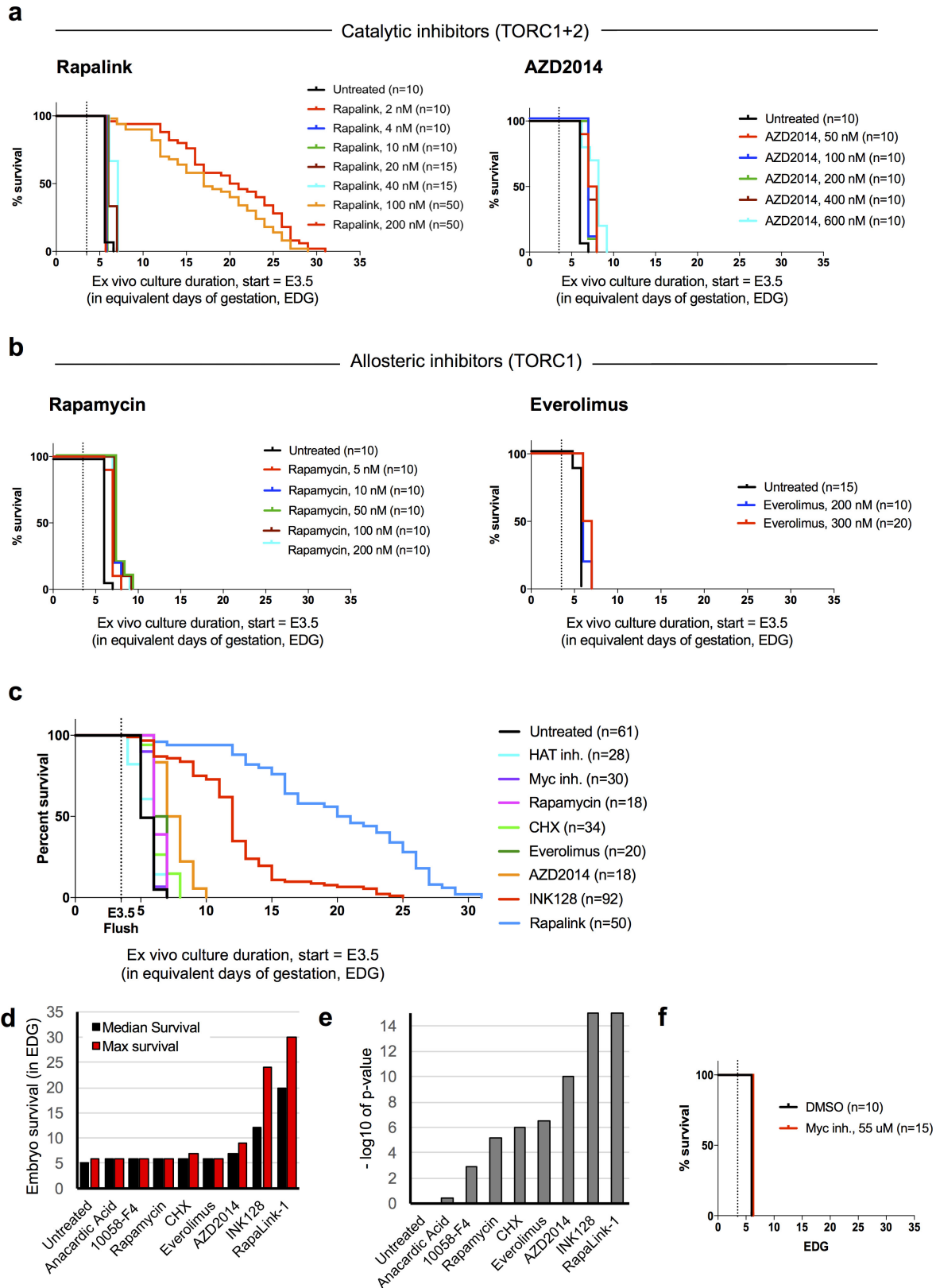
Pairwise GO term expression analysis. For each of the 3,772 gene ontology terms that are associated with at least 10 genes³⁴, we defined the gene ontology term expression as the mean FPKM values of genes associated with the corresponding term. In Fig. 4f, the \log_2 fold-change of gene ontology term expressions between paused ES cells and serum ES cells was plotted on the y axis against that between various samples in ref. 20 and E4.5 EPI on the x axis. The Spearman correlation coefficient of the 3,772 gene ontology terms is indicated. Extended Data Figure 10a was generated similarly, but with the \log_2 fold-change of gene ontology term expressions between *Myc* DKO and wild-type cells from ref. 6 on the y axis.

Pairwise pathway expression analysis. For each of the 281 KEGG pathways that contain at least 10 genes³⁵, we defined the pathway expression as the mean FPKM values of genes associated with the corresponding pathway. In Extended Data Fig. 9b, the \log_2 fold change of pathway expressions between paused ES cells and serum ES cells was plotted on the y axis against that between various samples in ref. 20 and E4.5 EPI. The Spearman correlation coefficient of the 281 pathways was indicated. Extended Data Fig. 10c was generated similarly, but with the \log_2 fold change of pathway expressions between *Myc* dKO and wild-type cells from ref. 6 on the y axis.

Code availability. Custom codes used for the RNA-seq analysis are available upon request.

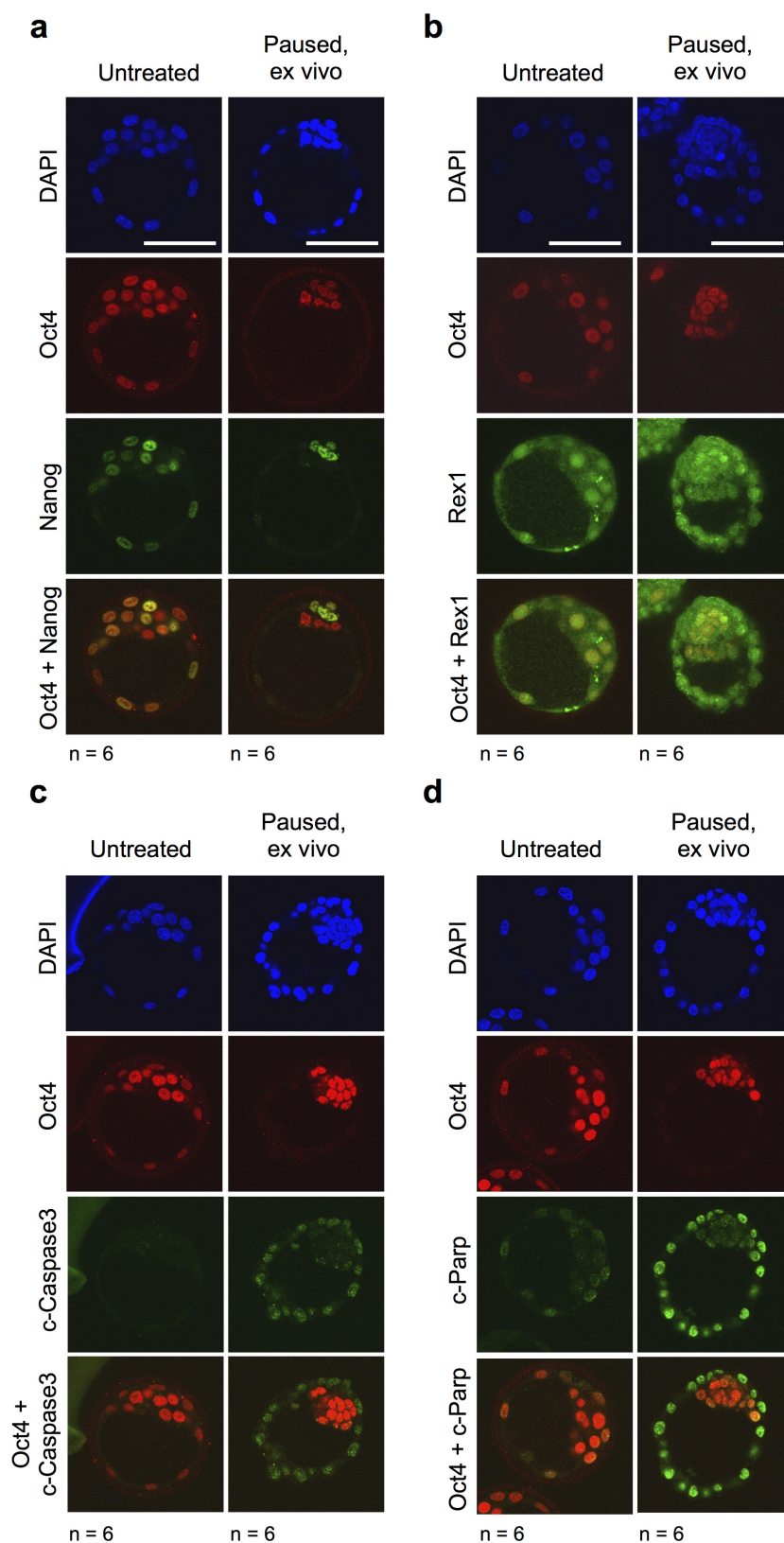
Data availability. RNA-seq data have been deposited in Gene Expression Omnibus (GEO) under accession number GSE81285. RNA-seq data from refs 6 and 20 are available under the accession numbers GSE74337 and E-MTAB-2958. The authors declare that all other data supporting the findings of this study are available within the paper and its supplementary information files.

25. Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996).
26. Behringer, R., Gertsenstein, M., Nagy, K. V. & Nagy, A. *Manipulating the Mouse Embryo*. (2013).
27. Lee, K.-H., Chuang, C.-K., Guo, S.-F. & Tu, C.-F. Simple and efficient derivation of mouse embryonic stem cell lines using differentiation inhibitors or proliferation stimulators. *Stem Cells Dev.* **21**, 373–383 (2012).
28. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
29. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
30. George, S. H. L. *et al.* Developmental and adult phenotyping directly from mutant embryonic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 4455–4460 (2007).
31. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
32. External RNA Controls Consortium. Proposed methods for testing and selecting the ERCC external RNA controls. *BMC Genomics* **6**, 150 (2005).
33. Lazar, C. *et al.* Batch effect removal methods for microarray gene expression data integration: a survey. *Brief. Bioinform.* **14**, 469–490 (2013).
34. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
35. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
36. Climent, M. *et al.* Functional analysis of Rex1 during preimplantation development. *Stem Cells Dev.* **22**, 459–472 (2013).
37. Aziz, M. & Alexandre, H. The origin of the nascent blastocoele in preimplantation mouse embryos ultrastructural cytochemistry and effect of chloroquine. *Roux Arch. Dev. Biol.* **200**, 77–85 (1991).



Extended Data Figure 1 | Catalytic mTOR inhibitors can induce blastocyst pausing. **a, b**, Kaplan–Meier survival curves of blastocysts cultured with two catalytic mTOR inhibitors Rapalink-1 and AZD2014 (**a**) and two allosteric mTOR inhibitors rapamycin and everolimus (**b**) at different concentrations. **c**, Kaplan–Meier survival curves of blastocysts

cultured with all 5 mTOR inhibitors used in this study. The concentrations yielding the best survival outcome are shown. **d**, Kaplan–Meier survival curves of blastocysts cultured with the Myc-inhibitor 10058-F4 at 55 μ M, as in ref. 6. Dotted lines indicate the time of blastocyst flushing (E3.5).

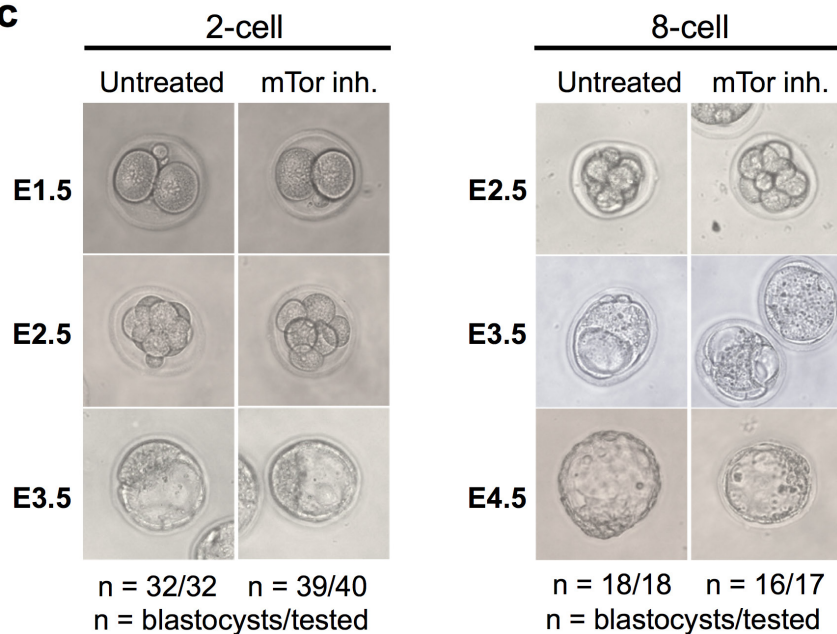


Extended Data Figure 2 | The *ex vivo* paused epiblast retains naive pluripotency markers and is devoid of apoptosis. It is possible that the signs of apoptosis in the trophoectoderm contribute to the eventual demise of mTOR-inhibited blastocysts after prolonged culture. **a–d**, Immunofluorescence images of untreated versus paused blastocysts

for Nanog (a), Rex1 (b), cleaved caspase 3 (c) and cleaved Parp1 (d). Oct4 staining is shown in all cases as a marker of the ICM. Note that the ubiquitous staining pattern for Rex1 is as expected³⁶. Scale bar, 50 μ m. The number of embryos analysed (*n*) is indicated.

a

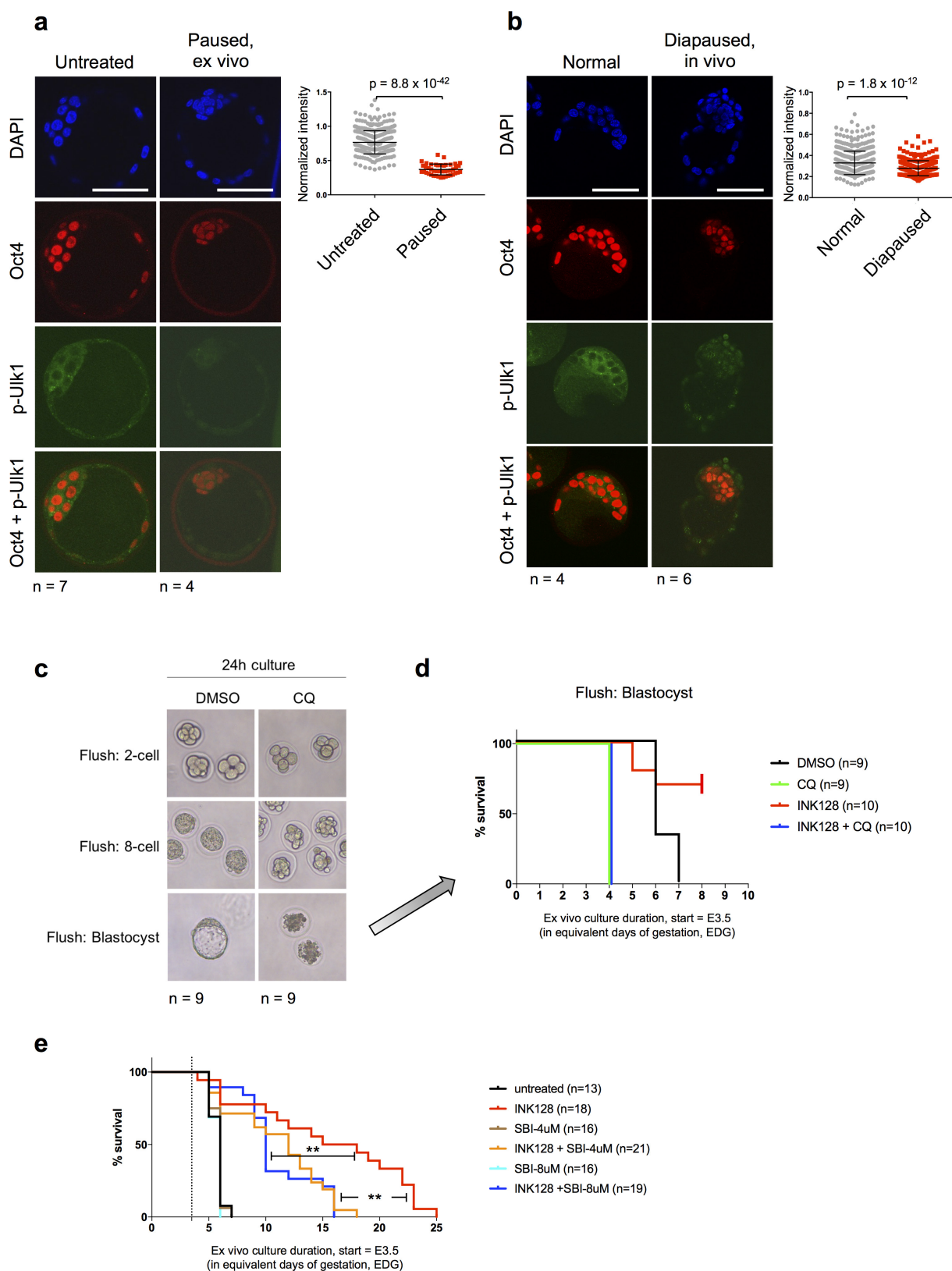
		# Embryos transferred	Pause duration (days)	# (%) Implantations	# (%) Fetuses	Delivery
Controls (Pooled data)	Non-surgical embryo transfer (NSET)	57	-	25 (44%)	11 (19%)	Sacrificed at E10.5
	Surgical embryo transfer (ET)	46	-	20 (43%)	10 (22%)	Sacrificed at E10.5
		# Embryos transferred	Pause duration (days)	# (%) Resorptions	# (%) Live birth	Delivery
Paused (Data from individual transfers)	Transfer 1 (UCSF, NSET)	20	5	5 (25%)	2 (10%)	Cesarean
	Transfer 2 (UCSF, NSET)	10	5	5 (50%)	1 (10%)	Cesarean
	Transfer 3 (UCSF, NSET)	19	5	2 (10%) Carcasses	2 (10%)	Natural
	Transfer 4 (UCSF, NSET)	13	7	N/A	0	N/A
	Transfer 5 (UCSF, NSET)	9	5	N/A	3 (33%)	Natural
	Transfer 6 (TCP, ET)	8	4	N/A	5 (62.5%)	Natural
	Transfer 7 (TCP, ET)	10	4	N/A	2 (20%)	Natural
	Transfer 8 (TCP, ET)	11	4	N/A	0	N/A

b**c**

Extended Data Figure 3 | Only blastocyst-stage embryos can be sustainably paused and they can give rise to live, fertile mice.

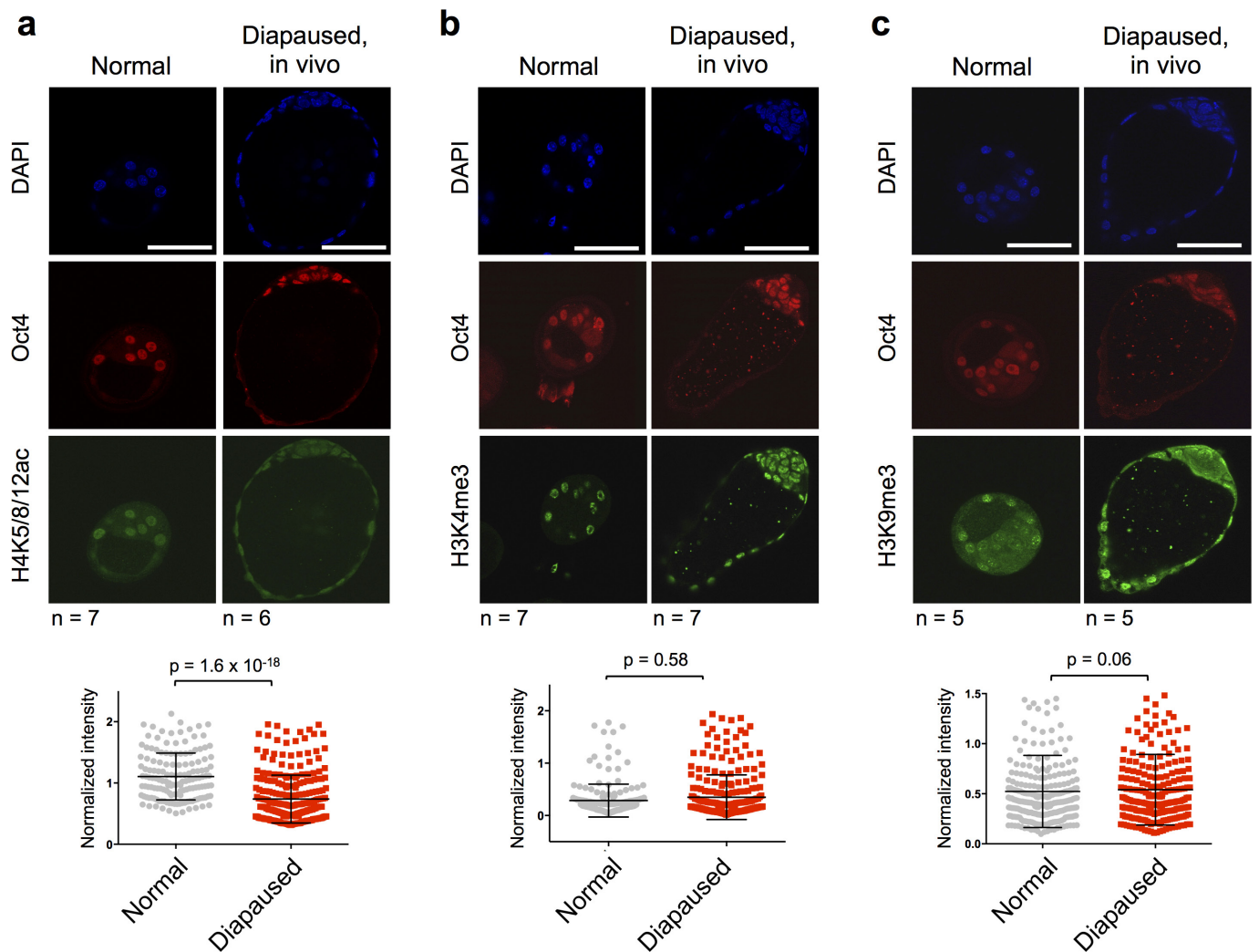
a. Detailed information on the generation of live mice from paused blastocysts by transfer into pseudo-pregnant surrogate females. Conditions used and live birth events are indicated. NSET, non-surgical embryo transfer; ET, embryo transfer (surgical). **b.** Live pups born as a result of mating mice generated from paused blastocysts with wild-type ICR mice.

All tested mice (5 out of 5, from a total of 15 mice generated from paused blastocysts) proved to be fertile. **c.** Representative images of *ex vivo* cultured 2- or 8-cell embryos with or without the mTOR inhibitor. Cleavage-stage embryos proceed to generate blastocysts even in presence of mTOR inhibitor. *n* represents number of blastocysts developed per number of tested cleavage stage embryos.



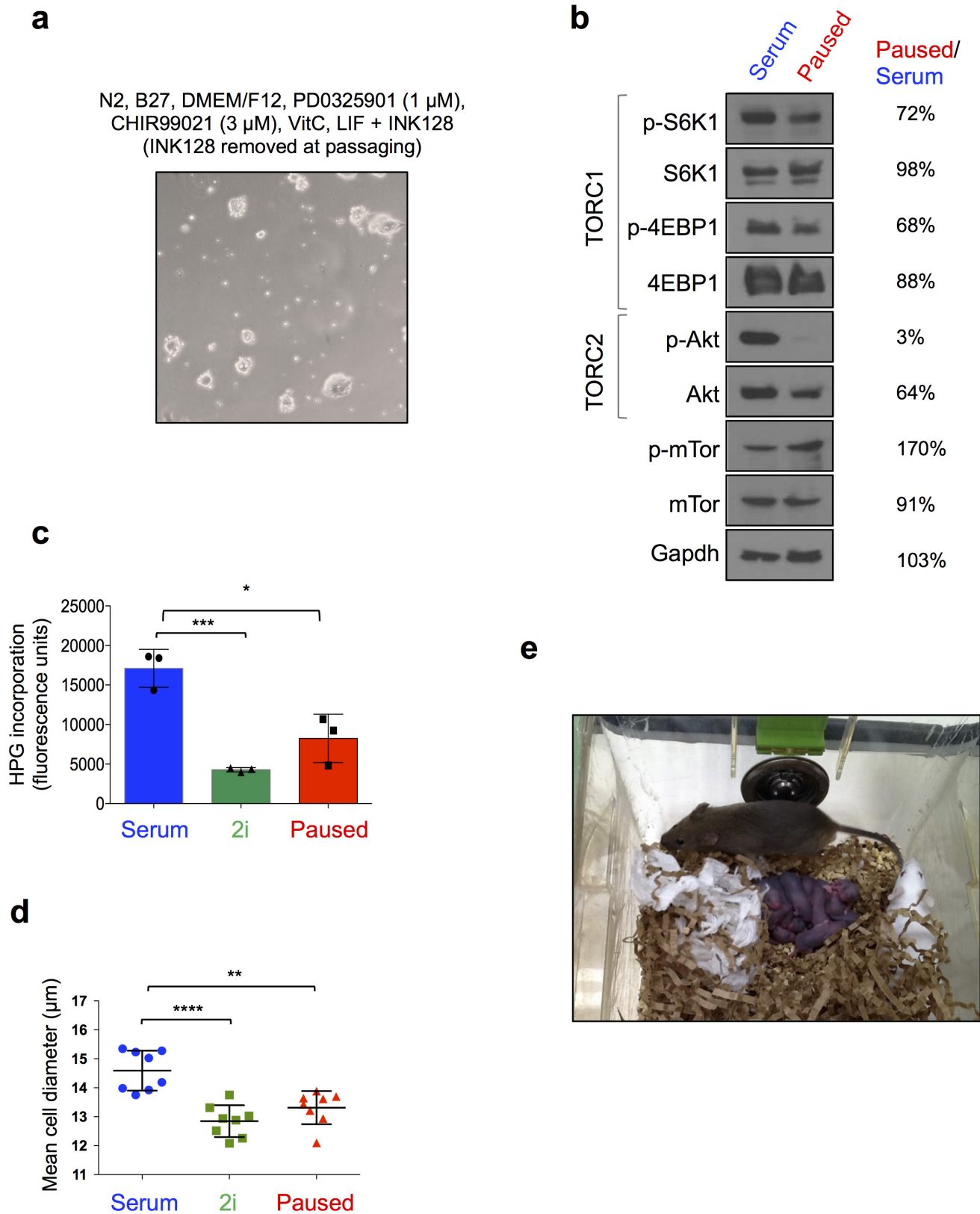
Extended Data Figure 4 | Paused blastocysts show signs of autophagy. **a, b**, Immunofluorescence images of phosphorylated Ulk1 in untreated versus paused (**a**), or normal versus diapaused (**b**) embryos. Oct4 staining is shown as a marker of the ICM. Scale bar, 50 μ m. The number of embryos analysed (n) is indicated. Error bars indicate standard deviation. P values are from two-tailed unpaired Student's t -test. **c**, Images showing embryos retrieved at different developmental stages incubated with the autophagy/lysosomal inhibitor chloroquine (CQ) for 24 h. The number of embryos analysed (n) is indicated. **d**, Kaplan–Meier survival curves of blastocysts

cultured with chloroquine in the presence or absence of the mTOR-inhibitor INK128. CQ treatment leads to blastocysts collapse, as previously reported³⁷. The experiment was terminated at EDG8.5. **e**, Kaplan–Meier survival curves of blastocysts cultured with the autophagy/Ulk1 inhibitor SBI-0206965 in the presence or absence of the mTOR inhibitor INK128. Inhibition of Ulk1 leads to decreased survival of paused blastocysts. $^{**}P < 0.01$ from log-rank (Mantel–Cox) test. Dotted lines indicate the time of blastocyst flushing (E3.5).



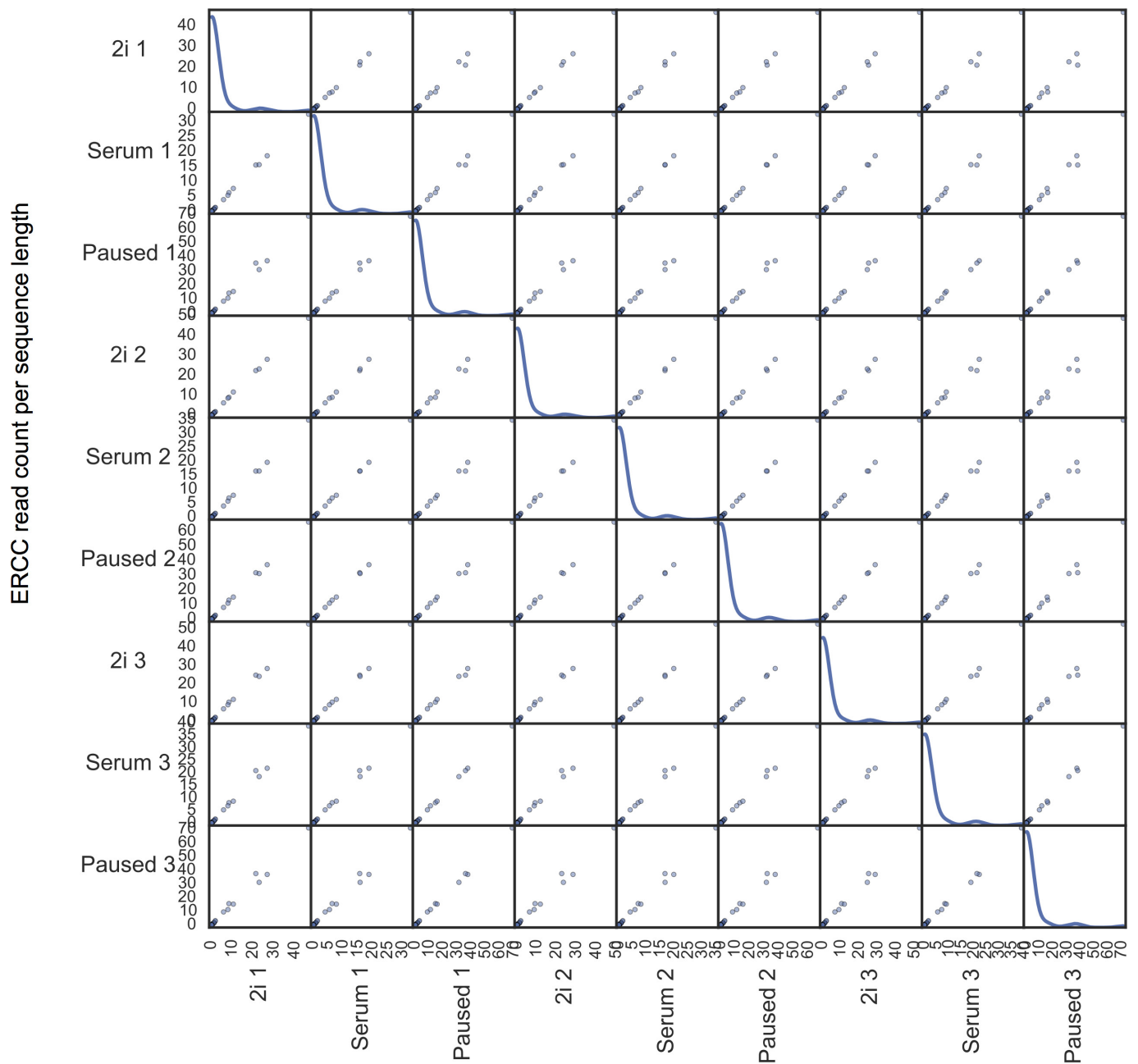
Extended Data Figure 5 | *In vivo* diapaused blastocysts have reduced histone H4 acetylation. **a–c,** Immunofluorescence images and quantification at the single ICM cell level of normal versus *in vivo* diapaused blastocysts for H4K5/8/12ac (**a**), H3K4me3 (**b**) and H3K9me3

(**c**) levels. Oct4 staining is shown in all cases as a marker of the ICM. Scale bar, 50 μ m. The number of embryos analysed (*n*) is indicated. Graphs show data pooled from all embryos. Error bars indicate standard deviation. *P* values are from two-tailed unpaired Student's *t*-test.



Extended Data Figure 6 | Partial inhibition of mTOR activity results in ES cell pausing. **a**, Representative image showing morphology of ES cells paused in 2i medium, with removal of the mTOR inhibitor at passaging. Under these conditions, ES cells can be discontinuously paused in 2i medium. **b**, Western blot showing moderately reduced 4EBP1 and S6K1 phosphorylation, which are mediated by mTORC1, and abolished Akt phosphorylation, which is mediated by mTORC2, in paused ES cells. For gel source data, see Supplementary Fig. 1. **c**, Flow cytometry analysis of

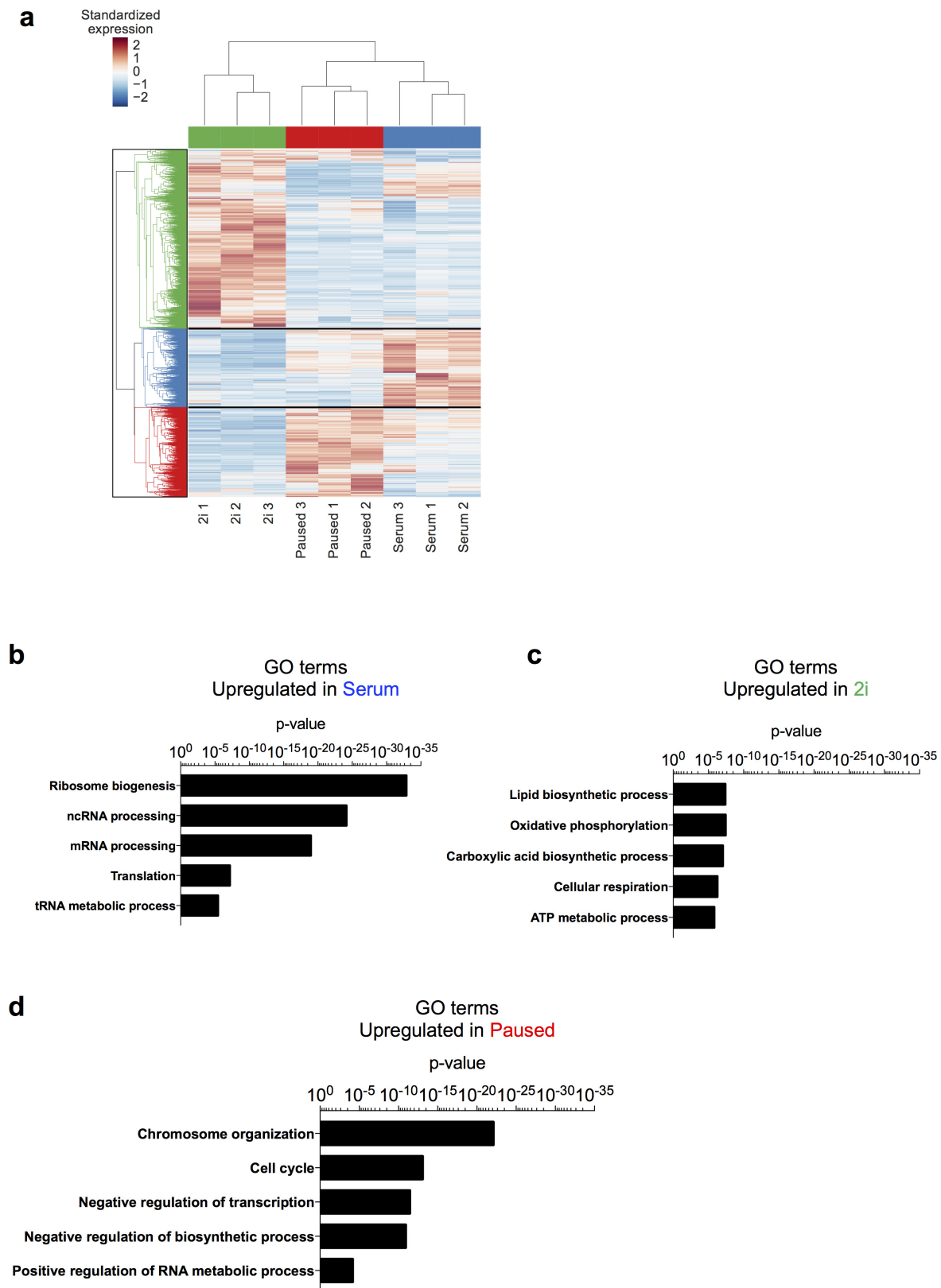
nascent translation in the three states measured by HPG incorporation in E14 cells, in triplicates. **d**, Analysis of cell diameter in the three states, with each data point representing a population average of at least 300 cells per measurement. Values are represented as mean \pm s.d. *P* values are from two-tailed unpaired Student's *t*-test. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, NS, not significant. **e**, Live pups born as a result of mating highly chimeric males generated using paused G4 ES cells with wild-type CD-1(ICR) female mice.



ERCC read count per sequence length

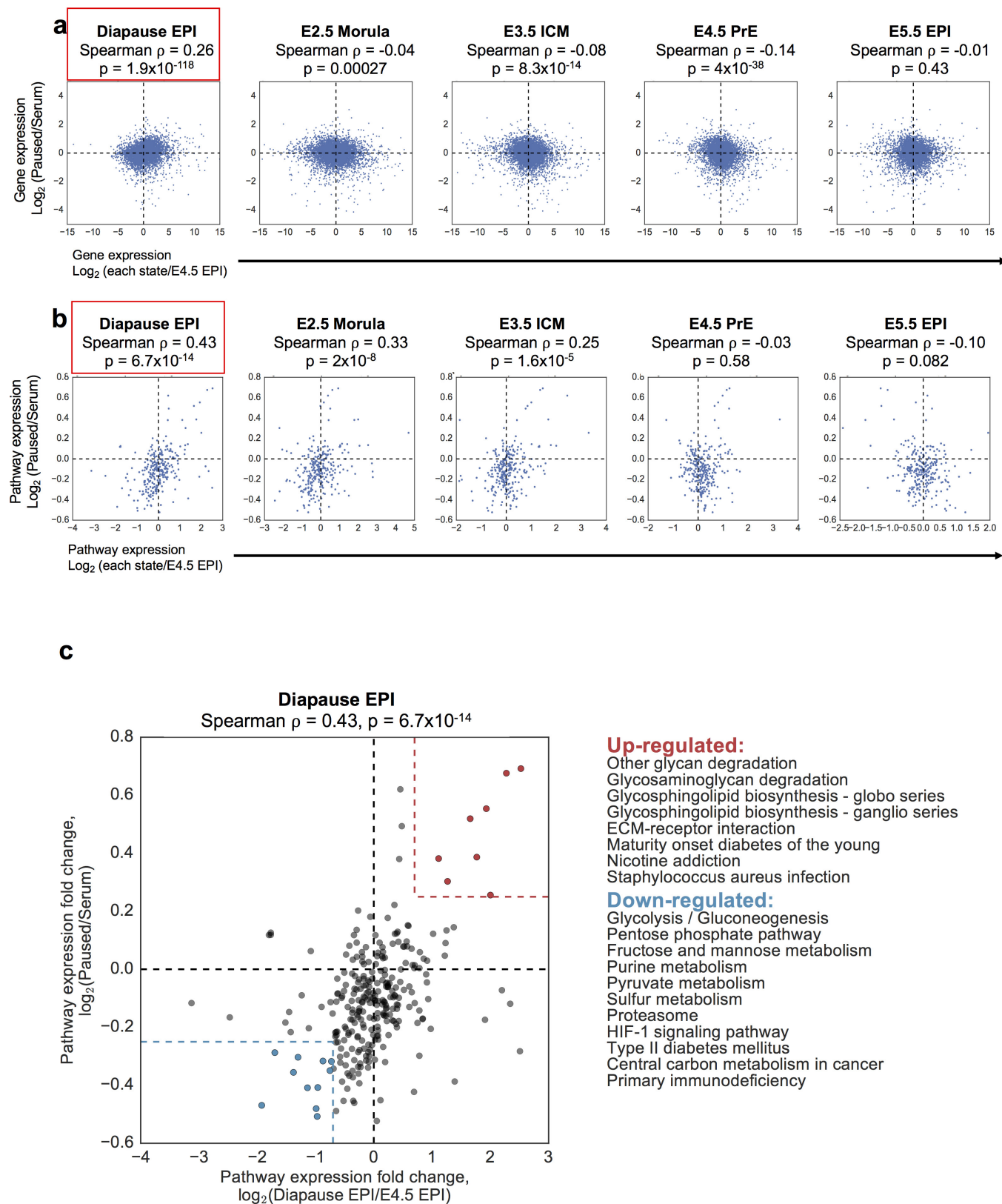
Extended Data Figure 7 | ERCC sequence abundances followed a highly linear trend in all pairs of samples across at least five orders of magnitude. The absolute abundance of mRNA transcripts was estimated using the ERCC92 RNA spike-ins³². Sequencing reads were aligned to the 92 reference spike-in sequences and the abundance of these sequences between different samples was compared. ERCC sequence abundances

followed a highly linear trend in all pairs of samples across at least 5 orders of magnitude (Pearson correlation coefficient larger than 99.7%). Diagonal entries in the figure matrix show the density of read counts. ERCC RNAs are consistently detected at higher abundance in the paused state compared to serum and 2i.



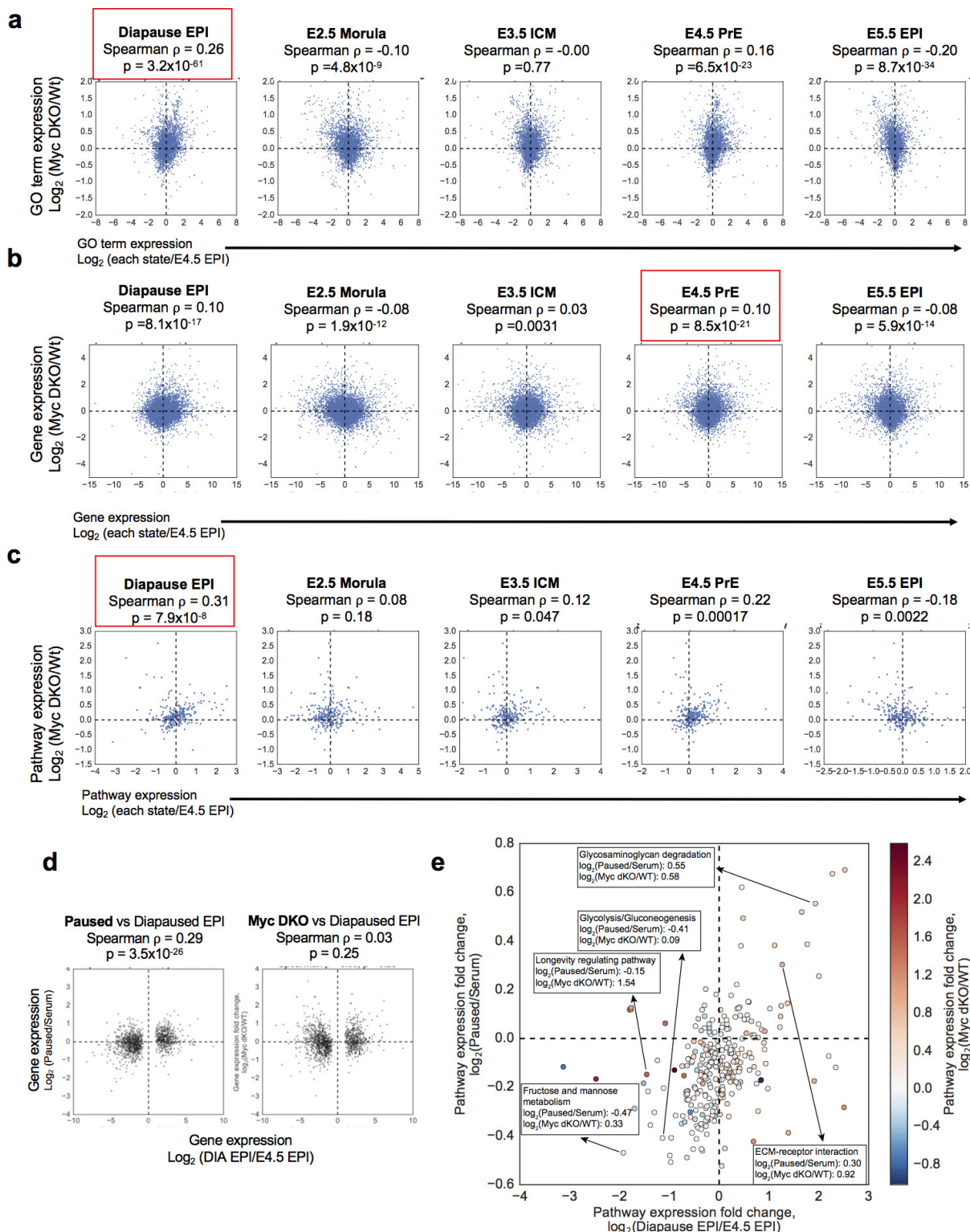
Extended Data Figure 8 | Distinct functional annotations are associated with different states of ES cells. **a**, Clustering of dynamically expressed genes. Heat map shows 3,864 dynamically expressed genes (differentially expressed between any two of the 2i, serum, and paused states and mean FPKM > 10 in at least one state). The FPKM value of each gene was standardized across the nine samples by subtracting the mean and then

dividing by the standard deviation. Hierarchical clustering was performed using the standardized expression values using Euclidean metric and average linkage. **b–d**, Selected GO terms enriched in the annotations of genes upregulated in serum (**b**), 2i (**c**) or paused (**d**) ES cells. See Supplementary Table 3 for complete list of significant gene ontology terms associated with each ES cell state.



Extended Data Figure 9 | Analysis of RNA-seq data indicates that paused ES cells mimic embryonic diapause. **a, b,** Scatter plots showing gene expression (9,418 genes that have FPKM >10 in any of our or Boroviak's samples are shown) (**a**) and pathway expression (**b**) comparing paused ES cells to different developmental stages. Spearman correlation coefficients and P values indicate a significant similarity of paused ES cells to the diapaused epiblast. y axes represent \log_2 fold change in gene expression (**a**) or pathway expression (**b**) in paused versus serum ES

cells. x axes represent \log_2 fold changes in gene expression (**a**) or pathway expression (**b**) in the different developmental stages (indicated above plots) versus E4.5 epiblast. Red boxes indicate the developmental stage that paused ES cells are closest to in each analysis. **c,** Scatter plot showing pathway expression in paused/serum ES cells (y axis) versus diapaused epiblast/E4.5 epiblast (x axis) as in (**b**). Pathways coordinately upregulated (red) or downregulated (blue) in paused ES cells and diapaused epiblast are indicated next to the scatter plot.



Extended Data Figure 10 | Analysis of RNA-seq data indicates that *Myc* double knockout (DKO) cells have similarities to embryonic diapause at the GO term and pathway levels, but to a lesser extent than paused ES cells. a–c, Scatter plots showing gene ontology term expression (a), gene expression (b), and pathway expression (c) comparing *Myc* DKO ES cells⁶ to different developmental stages, similarly to Fig. 4f and Extended Data Fig. 9a, b. Red boxes indicate the developmental stage that *Myc* DKO ES cells are closest to in each analysis. d, Scatter plots showing ‘diapause-driver’ gene expression (1,324 genes that are differentially expressed between the diapause epiblast and E4.5 epiblast) in paused ES cells (left panel) and *Myc* DKO cells (right panel) compared to the

diapause epiblast. Pseudocount 1 was used when calculating the log₂ fold changes. Spearman correlation coefficient and *P* value indicates a statistically significant similarity of paused ES cells, but not *Myc* DKO cells, to the diapause epiblast. e, Scatter plot showing pathway expression in paused ES cells versus the diapause epiblast, as in Extended Data Fig. 9c. Colour-coded expression levels of these pathways in *Myc* DKO ES cells are superimposed onto this graph. Numerical values for selected pathways are indicated. Upregulated pathways are in general concordant in paused and *Myc* DKO ES cells, whereas downregulated pathways are mostly discordant.

RIPK1 counteracts ZBP1-mediated necroptosis to inhibit inflammation

Juan Lin^{1*}, Snehlata Kumari^{1*}, Chun Kim^{1*}, Trieu-My Van¹, Laurens Wachsmuth¹, Apostolos Polykratis¹ & Manolis Pasparakis¹

Receptor-interacting protein kinase 1 (RIPK1) regulates cell death and inflammation through kinase-dependent and -independent functions^{1–7}. RIPK1 kinase activity induces caspase-8-dependent apoptosis and RIPK3 and mixed lineage kinase like (MLKL)-dependent necroptosis^{8–13}. In addition, RIPK1 inhibits apoptosis and necroptosis through kinase-independent functions, which are important for late embryonic development and the prevention of inflammation in epithelial barriers^{14–18}. The mechanism by which RIPK1 counteracts RIPK3–MLKL-mediated necroptosis has remained unknown. Here we show that RIPK1 prevents skin inflammation by inhibiting activation of RIPK3–MLKL-dependent necroptosis mediated by Z-DNA binding protein 1 (ZBP1, also known as DAI or DLM1). ZBP1 deficiency inhibited keratinocyte necroptosis and skin inflammation in mice with epidermis-specific RIPK1 knockout. Moreover, mutation of the conserved RIP homotypic interaction motif (RHIM) of endogenous mouse RIPK1 (RIPK1^{mRHIM}) caused perinatal lethality that was prevented by RIPK3, MLKL or ZBP1 deficiency. Furthermore, mice expressing only RIPK1^{mRHIM} in keratinocytes developed skin inflammation that was abrogated by MLKL or ZBP1 deficiency. Mechanistically, ZBP1 interacted strongly with phosphorylated RIPK3 in cells expressing RIPK1^{mRHIM}, suggesting that the RIPK1 RHIM prevents ZBP1 from binding and activating RIPK3. Collectively, these results show that RIPK1 prevents perinatal death as well as skin inflammation in adult mice by inhibiting ZBP1-induced necroptosis. Furthermore, these findings identify ZBP1 as a critical mediator of inflammation beyond its previously known role in antiviral defence and suggest that ZBP1 might be implicated in the pathogenesis of necroptosis-associated inflammatory diseases.

Mice with epidermis-specific RIPK1 deficiency (*Ripk1*^{fl/fl}*K14-cre*^{Tg/WT}, hereafter referred to as RIPK1^{E-KO}) develop skin inflammation owing to RIPK3–MLKL-dependent keratinocyte necroptosis¹⁴. We hypothesized that other RHIM-containing proteins may induce RIPK3–MLKL-mediated necroptosis in RIPK1-deficient keratinocytes. In addition to RIPK1 and RIPK3, the only other proteins containing RHIM in humans and mice are TRIF and ZBP1 (refs 3, 6, 19, 20). We showed previously that TRIF deficiency very mildly ameliorates but does not prevent skin inflammation in RIPK1^{E-KO} mice¹⁴, suggesting that TRIF is not essential for RIPK3 activation and necroptosis in RIPK1-deficient epidermal keratinocytes. ZBP1 is a RHIM-containing protein previously identified as a cytoplasmic DNA sensor capable of inducing type I interferon expression and NF- κ B activation^{19–23}. More recently, ZBP1 was shown to induce necroptosis by activating RIPK3 independently of RIPK1 in response to cytomegalovirus infection²⁴. We therefore hypothesized that ZBP1 might be implicated in triggering RIPK3 activation and necroptosis in epidermal keratinocytes of RIPK1^{E-KO} mice. Immunoblot analysis of epidermal extracts showed that ZBP1 was expressed at low levels in wild-type mice but its expression was strongly increased in the epidermis of RIPK1^{E-KO} mice at the

age of four weeks (Fig. 1a), supporting the idea that ZBP1 could be involved in triggering keratinocyte necroptosis in this model.

To address the potential role of ZBP1 in triggering keratinocyte necroptosis and skin inflammation in RIPK1^{E-KO} mice, we crossed them with *Zbp1*^{-/-} mice²⁵. RIPK1^{E-KO}*Zbp1*^{-/-} mice did not show macroscopic signs of skin disease at the age of four weeks, in contrast to RIPK1^{E-KO} animals that displayed inflammatory skin lesions at this age (Extended Data Fig. 1a). Histological analysis confirmed that 4–5-week-old RIPK1^{E-KO}*Zbp1*^{-/-} animals did not develop skin

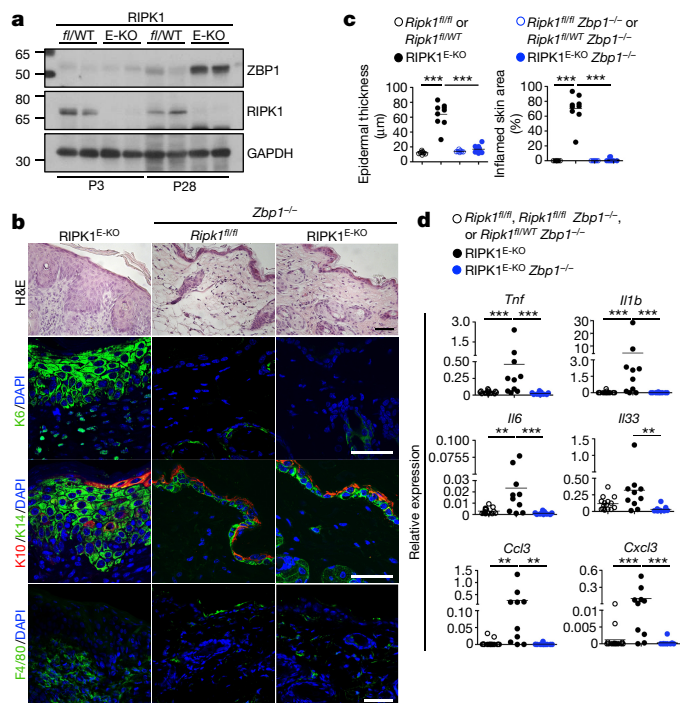


Figure 1 | ZBP1 induces keratinocyte necroptosis and skin inflammation in RIPK1^{E-KO} mice. **a**, Immunoblot analysis of ZBP1, RIPK1 and GAPDH in epidermal protein extracts from wild type (*Ripk1*^{fl/WT}) and RIPK1^{E-KO} mice at postnatal day 3 (P3) and P28. Lanes represent samples from individual mice. For gel source data, see Supplementary Fig. 1. **b**, Skin sections from 4–5-week-old mice were stained with H&E or immunostained with the indicated antibodies. Representative images shown (RIPK1^{E-KO} *n* = 9 for H&E and *n* ≥ 6 for immunostainings; RIPK1^{E-KO}*Zbp1*^{-/-} *n* = 10 for H&E and *n* ≥ 3 for immunostainings). Scale bars, 50 μ m. **c**, Microscopic quantification of epidermal thickness and inflamed skin area in 4–5-week-old mice with the indicated genotypes. **d**, qRT-PCR analysis of the mRNA expression of the indicated cytokines and chemokines in RNA isolated from total skin from 4–5-week-old mice with the indicated genotypes. **P* ≤ 0.05; ***P* ≤ 0.01; ****P* ≤ 0.005.

¹Institute for Genetics, Centre for Molecular Medicine (CMC), and Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, 50931 Cologne, Germany.

*These authors contributed equally to this work.

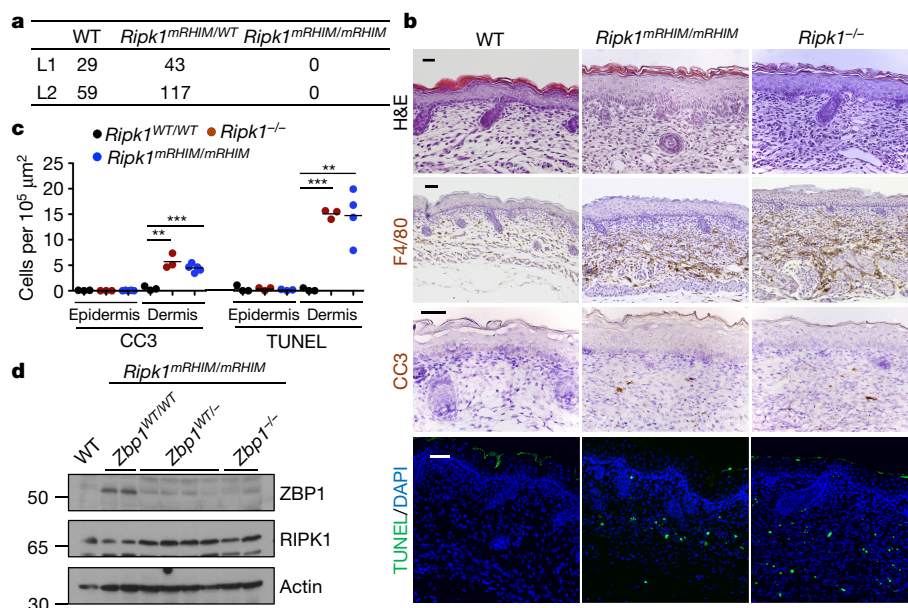


Figure 2 | Mutation of the RIPK1 RHIM domain causes perinatal lethality and inflammatory skin hyperplasia in mice. **a**, Table showing the numbers of weaned offspring of *Ripk1^{mRHIM/WT}* parents from two independently generated knock-in lines (L1 & L2). **b**, Skin sections from E18.5 pups were stained with H&E or TUNEL or immunostained for CC3 or F4/80. Representative images shown (wild type $n = 6$ for H&E, $n = 3$ for TUNEL, anti-CC3 and anti-F4/80; *Ripk1^{mRHIM/mRHIM}* $n = 5$ for H&E and

anti-CC3, $n = 4$ for TUNEL and $n = 3$ for anti-F4/80; *Ripk1^{-/-}* $n = 3$ for H&E, TUNEL, anti-F4/80 and anti-CC3). Scale bars, 50 μm.

c, Microscopic quantification of CC3⁺ and TUNEL⁺ cells on skin sections from E18.5 pups with the indicated genotypes. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.005$. **d**, Immunoblot analysis of total skin lysates from E18.5 pups of the indicated genotypes. Lanes represent samples from individual embryos. For gel source data, see Supplementary Fig. 1.

lesions, as shown by normal epidermal thickness and typical expression of epidermal differentiation markers including keratins 14, 10 and 6 (Fig. 1b, c). In addition, the skin of *RIPK1^{E-KO} Zbp1^{-/-}* mice did not show increased infiltration of F4/80⁺ myeloid cells and upregulation of inflammatory cytokines and chemokines (Fig. 1b, d). ZBP1 deficiency also reduced the number of keratinocytes stained with terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) in the epidermis of *RIPK1^{E-KO}* mice (Extended Data Fig. 1e, f), suggesting that it prevented keratinocyte necroptosis. *RIPK1^{E-KO} Zbp1^{-/-}* mice remained healthy until the age of 18–20 weeks, but subsequently progressively developed inflammatory skin lesions; however, these remained milder and focal compared to the severe inflammation affecting the entire skin observed in 4–5 week-old *RIPK1^{E-KO}* mice (Extended Data Fig. 1b–d). Therefore, ZBP1 deficiency strongly inhibited, but did not completely prevent, skin inflammation in *RIPK1^{E-KO}* mice, in contrast to *RIPK3* or *MLKL* deficiency that fully abrogated lesion development in these animals¹⁴. These results suggest that ZBP1 plays a critical role in the induction of *RIPK3*–*MLKL*-dependent keratinocyte necroptosis in *RIPK1^{E-KO}* mice but, in its absence, alternative mechanisms can activate *RIPK3* to trigger necroptosis. Although keratinocyte-specific *TRIF* deficiency did not considerably inhibit skin inflammation in *RIPK1^{E-KO}* mice¹⁴, it is possible that *TRIF* might contribute to skin lesion development in the absence of ZBP1 in adult *RIPK1^{E-KO} Zbp1^{-/-}* animals. To address whether ZBP1 is generally required for keratinocyte necroptosis we assessed the role of ZBP1 in mice with epidermis-specific *FADD* deficiency (*FADD^{E-KO}*), which develop skin inflammation owing to *RIPK3*-mediated keratinocyte necroptosis²⁶. *FADD^{E-KO} Zbp1^{-/-}* mice developed skin inflammation similarly to *FADD^{E-KO}* mice, showing that ZBP1 deficiency did not inhibit *RIPK3*-dependent skin inflammation in this model (Extended Data Fig. 2). Thus, ZBP1 induces keratinocyte necroptosis in the absence of *RIPK1* but it is not required for keratinocyte necroptosis and skin inflammation caused by epidermal *FADD* deficiency.

We reasoned that *RIPK1* might prevent ZBP1-mediated *RIPK3* activation by interacting with these two proteins through its RHIM domain. To specifically address the role of the RHIM domain of *RIPK1*

in vivo, we generated knock-in mice expressing a mutated *RIPK1* protein in which the QIG conserved amino acids of the RHIM domain at position 529–531 were substituted with alanines (*RIPK1*(QIG529→531AAA), hereafter referred to as *RIPK1^{mRHIM}*) using CRISPR/Cas9-mediated gene targeting in mouse zygotes (Extended Data Fig. 3a). Genotyping of progeny obtained from intercrossing heterozygous *Ripk1^{mRHIM/WT}* mice from two independently generated knock-in lines failed to identify any homozygous *Ripk1^{mRHIM/mRHIM}* mice at weaning age (Fig. 2a). Examination of pups obtained from timed matings revealed that *Ripk1^{mRHIM/mRHIM}* mice were alive at E18.5 but died perinatally, similarly to *Ripk1^{-/-}* mice^{15–17,27}. Histological analysis of E18.5 pups revealed epidermal hyperplasia, massive infiltration of F4/80⁺ myeloid cells and increased number of TUNEL⁺ and a few cleaved caspase-3 positive (CC3⁺) cells in the dermis of *Ripk1^{mRHIM/mRHIM}* and *Ripk1^{-/-}* pups (Fig. 2b, c). Analysis of intestinal tissues revealed scarce presence of CC3⁺ apoptotic cells in *Ripk1^{mRHIM/mRHIM}* pups compared to the increased number of CC3⁺ cells found in the intestine of *Ripk1^{-/-}* mice (Extended Data Fig. 3b, c). Earlier studies showed that the epidermal hyperplasia observed in *Ripk1^{-/-}* pups depends on *RIPK3*–*MLKL*-mediated necroptosis¹⁷, whereas the intestinal epithelial cell apoptosis is driven by *FADD*/caspase-8-dependent apoptosis^{14,15,17,18}. Inhibition of both necroptosis and apoptosis was required to overcome the perinatal lethality of *Ripk1^{-/-}* mice^{15–17}. However, *RIPK3* deficiency was sufficient to prevent perinatal lethality of *Ripk1^{mRHIM/mRHIM}* mice, as *Ripk1^{mRHIM/mRHIM} Ripk3^{-/-}* animals reached adulthood without showing any signs of pathology, at least up to the age of 6 months (Extended Data Table 1, Extended Data Fig. 4a, b). Interestingly, *Ripk1^{mRHIM/mRHIM} Ripk3^{WT/-}* mice survived to adulthood, suggesting that reduction of *RIPK3* protein levels by about 50% was sufficient to prevent necroptosis and perinatal lethality in these animals (Extended Data Table 1). In addition, crossing with *MLKL*-deficient mice (Extended Data Fig. 3d) also rescued perinatal death of *Ripk1^{mRHIM/mRHIM}* animals, with *Ripk1^{mRHIM/mRHIM} Mlkl^{-/-}* mice surviving at least up to 4 months without signs of disease (Extended Data Table 1, Extended Data Fig. 4a). Furthermore, ZBP1 expression was upregulated in the skin of *Ripk1^{mRHIM/mRHIM}* pups

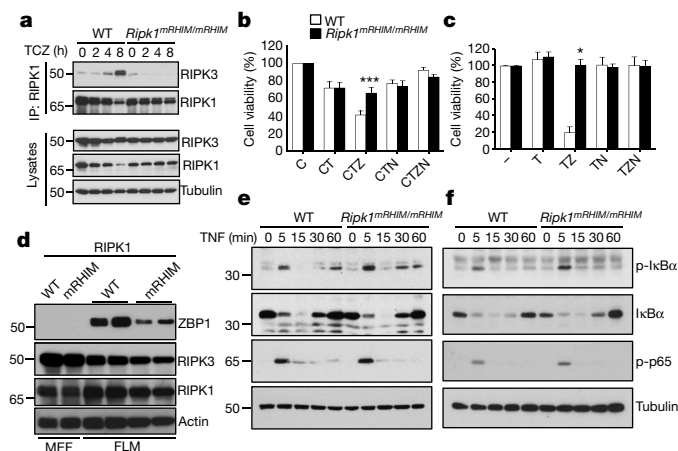


Figure 3 | Mutation of the RIPK1 RHIM domain prevents TNF-induced association of RIPK1 with RIPK3 and necroptosis. **a**, Immunoblot analysis with the indicated antibodies of RIPK1 immunoprecipitates and total lysates from primary wild-type and *Ripk1*^{mRHIM/mRHIM} MEFs treated with TNF (20 ng ml⁻¹), cycloheximide (1 μg ml⁻¹) and Z-VAD-FMK (20 μM) for the indicated periods of time. Representative data shown from two independent experiments. For gel source data, see Supplementary Fig. 1. **b, c**, Primary MEFs (**b**) or FLMs (**c**) from mice with the indicated genotypes were treated with combinations of TNF (20 ng ml⁻¹), cycloheximide (1 μg ml⁻¹), Z-VAD-FMK (20 μM) and Nec-1 (30 μM) for 18 h. Cell viability was determined by neutral red assay. Graphs show mean ± s.e.m. from pooled data from 5 (**b**) and 3 (**c**) independent experiments. **P* ≤ 0.05; ***P* ≤ 0.01; ****P* ≤ 0.005. **d**, Immunoblot analysis of primary MEFs or FLMs from wild-type or *Ripk1*^{mRHIM/mRHIM} mice with the indicated antibodies. Lanes represent primary cells from individual mice. For gel source data, see Supplementary Fig. 1. **e, f**, Primary MEFs (**e**) or FLMs (**f**) from wild-type or *Ripk1*^{mRHIM/mRHIM} mice were stimulated with TNF (20 ng ml⁻¹) for the indicated periods of time and NF-κB activation was assessed by immunoblotting with the indicated antibodies. Representative data shown from three independent experiments. For gel source data, see Supplementary Fig. 1.

(Fig. 2d) and ZBP1 deficiency also prevented perinatal lethality of these mice, with *Ripk1*^{mRHIM/mRHIM} *Zbp1*^{-/-} mice surviving at least up to the age of 5 months without showing apparent abnormalities (Extended Data Table 1, Extended Data Fig. 4a, b). TRIF knockout did not rescue the *Ripk1*^{mRHIM/mRHIM} mice (Extended Data Table 1). Therefore, in contrast to RIPK1 deficiency that causes perinatal lethality owing to both caspase-8-mediated apoptosis and RIPK3–MLKL-mediated necroptosis^{15–17}, mutation of RIPK1 RHIM caused perinatal death exclusively owing to ZBP1–RIPK3–MLKL-dependent necroptosis.

To confirm that the RIPK1(QIG→AAA) mutation disrupted the interaction of RIPK1 with RIPK3, we stimulated primary MEFs from *Ripk1*^{mRHIM/mRHIM} mice with TNF in the presence of cycloheximide and Z-VAD-FMK (TCZ treatment) for different periods of time to induce formation of the necrosome. Immunoblot analysis of RIPK1 immunoprecipitates revealed that RIPK3 strongly interacted with RIPK1 in TCZ-treated wild-type, but not in *Ripk1*^{mRHIM/mRHIM}, primary MEFs (Fig. 3a). Consistently, primary *Ripk1*^{mRHIM/mRHIM} MEFs showed reduced cell death in response to TCZ treatment and fetal liver macrophages (FLMs) from *Ripk1*^{mRHIM/mRHIM} pups were resistant to necroptosis induced by stimulation with TNF and Z-VAD-FMK (TZ treatment) (Fig. 3b, c). Notably, we routinely obtained reduced numbers of FLMs from *Ripk1*^{mRHIM/mRHIM} compared to wild-type embryos, and the expression levels of ZBP1 were reduced in the *Ripk1*^{mRHIM/mRHIM} cells (Fig. 3d), suggesting that *Ripk1*^{mRHIM/mRHIM} FLMs expressing high levels of ZBP1 may be counter-selected in these cultures. Therefore, disruption of the RHIM-dependent interaction of RIPK1 with RIPK3 protected primary FLMs and MEFs from TNF-induced necroptosis. TNF-induced NF-κB activation was not impaired in *Ripk1*^{mRHIM/mRHIM} MEFs or FLMs (Fig. 3e, f), showing that RHIM-dependent RIPK1

interactions are not required for TNFR1-induced proinflammatory signalling.

To address whether RIPK1 prevents keratinocyte necroptosis and skin inflammation in a RHIM-dependent manner, we crossed *Ripk1*^{mRHIM/WT} with *Ripk1*^{fl/WT} *K14-cre*^{Tg/WT} mice to generate *Ripk1*^{mRHIM/fl} *K14-cre*^{Tg/WT} mice (hereafter referred to as RIPK1^{mRHIM/E-KO}), which express exclusively the mutant RIPK1^{mRHIM} in keratinocytes. In contrast to *Ripk1*^{mRHIM/WT} mice that did not show pathology in their skin or other organs (data not shown), RIPK1^{mRHIM/E-KO} mice developed macroscopically visible signs of skin lesions starting at about 3–4 weeks after birth, which progressively developed to inflammatory skin disease by the age of 9–11 weeks (Fig. 4a–c, Extended Data Fig. 5a–d). Histological analysis showed that the skin lesions in RIPK1^{mRHIM/E-KO} mice resembled those seen in RIPK1^{E-KO} mice, characterized by epidermal hyperplasia and impaired differentiation, increased numbers of dying keratinocytes as well as increased F4/80⁺ myeloid cell infiltration and upregulation of inflammatory cytokine and chemokine expression (Fig. 4a–c, Extended Data Fig. 5c–d). As in RIPK1^{E-KO} mice, homozygous, but not heterozygous, MLKL deficiency prevented skin lesion development in RIPK1^{mRHIM/E-KO} mice at least up to the age of 22 weeks, showing that the inflammatory skin disease is triggered by MLKL-mediated keratinocyte necroptosis (Fig. 4a–c, Extended Data Fig. 5a–d). RIPK1^{mRHIM/E-KO} mice showed increased expression of ZBP1 in the skin, similarly to RIPK1^{E-KO} mice (Fig. 4d, e). ZBP1 was not expressed in primary keratinocytes from wild-type, RIPK1^{E-KO} or RIPK1^{mRHIM/E-KO} mice (Extended Data Fig. 6a), suggesting that its upregulation in the epidermis could be triggered by signals related to the *in vivo* tissue context. Indeed, the increased expression of *Ifnb1* in the skin of RIPK1^{mRHIM/E-KO} mice could be responsible for the upregulation of ZBP1 expression (Fig. 4e), as stimulation with IFNβ induced robust ZBP1 expression in cultured primary keratinocytes from wild-type, RIPK1^{E-KO} and RIPK1^{mRHIM/E-KO} mice (Extended Data Fig. 6b). In line with our findings in RIPK1^{E-KO} animals, ZBP1 deficiency prevented the development of skin lesions in RIPK1^{mRHIM/E-KO} mice at least up to the age of 21 weeks (Fig. 4a–c, Extended Data Fig. 5a, c, d). These results showed that RHIM-dependent RIPK1 function in epidermal keratinocytes is critical to prevent ZBP1-mediated activation of RIPK3–MLKL-driven necroptosis and skin inflammation.

We postulated that RIPK1 may bind ZBP1 and prevent its interaction with RIPK3. As ZBP1 is not expressed in MEFs, we transduced primary wild-type or *Ripk1*^{mRHIM/mRHIM} MEFs with a lentiviral vector expressing Flag-tagged murine ZBP1. Immunoblotting of anti-Flag immunoprecipitates with anti-RIPK3 antibodies showed that ZBP1 interacted weakly with RIPK3 in wild-type MEFs, but this interaction was strongly enhanced in *Ripk1*^{mRHIM/mRHIM} MEFs (Fig. 4f). A slower migrating RIPK3 species was detected in the anti-Flag immunoprecipitate suggesting that ZBP1 could preferentially associate with phosphorylated RIPK3. Indeed, immunoblotting with monoclonal antibodies specifically recognizing RIPK3 phosphorylated at serine 232 (ref. 28) revealed that the slower migrating band corresponded to phosphorylated RIPK3 (Fig. 4f). Caspase-8 or MLKL were not detected in the anti-Flag immunoprecipitate, suggesting that these proteins do not interact with ZBP1 under these conditions (Fig. 4f). Immunoblotting with anti-RIPK1 antibodies failed to detect RIPK1 in the anti-Flag immunoprecipitate (Fig. 4f). Moreover, reciprocal immunoprecipitation using anti-RIPK1 antibodies and immunoblotting with anti-Flag or anti-ZBP1 antibodies also failed to detect an interaction between RIPK1 and ZBP1 (Fig. 4g). Therefore, in contrast to RIPK3, RIPK1 did not interact with ZBP1 in primary MEFs. It is not clear why our results differ from previous studies showing that ZBP1 interacted with RIPK1 in 293T cells^{19,20}, but this could be related to the absence of RIPK3 expression in 293T cells⁹.

Taken together, our results showed that, in the absence of the RIPK1 RHIM domain, ZBP1 strongly interacted with RIPK3 and induced its autophosphorylation, which triggered downstream activation of MLKL and necroptosis. The inhibitory role of the RIPK1 RHIM domain is

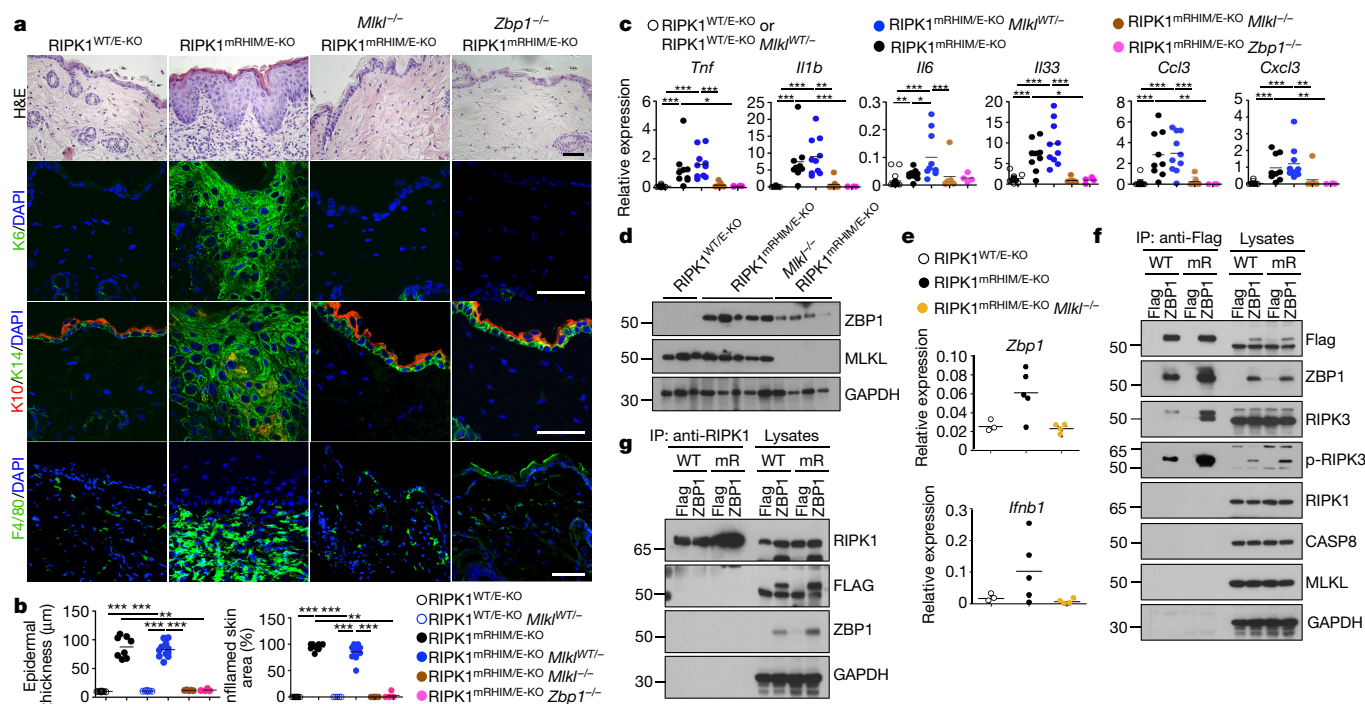


Figure 4 | RHIM-dependent RIPK1 function prevents MLKL-ZBP1-mediated necroptosis and skin inflammation. **a**, Skin sections from 9–11-week-old mice were stained with H&E or immunostained with the indicated antibodies. Representative images shown (RIPK1^{mRHIM/E-KO} *n* = 9 for H&E and *n* = 3 for immunostainings; RIPK1^{mRHIM/E-KO} *Mikl*^{-/-} *n* = 5 for H&E and *n* = 3 for immunostainings; RIPK1^{mRHIM/E-KO} *Zbp1*^{-/-} *n* = 4 for H&E and *n* = 3 for immunostainings). Scale bars, 50 μ m. **b**, **c**, Microscopic quantification of epidermal thickness and inflamed skin area (b) and qRT-PCR analysis of the mRNA expression of the indicated cytokines and chemokines (c) in 9–11-week-old mice with the indicated genotypes. **P* \leq 0.05; ***P* \leq 0.01; ****P* \leq 0.005.

particularly important for the maintenance of skin homeostasis during late embryonic life and in adult mice. As the lack of RIPK1, or mutation of its RHIM, specifically in the epidermis triggers keratinocytes necroptosis and inflammation starting a few weeks after birth, whereas ubiquitous RIPK1 deficiency or RHIM mutation triggers necroptosis of dermal cells and results in perinatal death, it is likely that the skin hyperplasia during late embryonic life and the associated perinatal lethality are caused by necroptosis of non-epithelial, perhaps stromal or myeloid, cells. Although the precise mechanism of the RIPK1 RHIM-dependent inhibition of ZBP1-mediated RIPK3 activation remains unknown at present, it is possible that RIPK1 associates with RIPK3 to prevent its interaction with ZBP1. At this stage, it is also unclear whether the nucleic acid sensing properties of ZBP1 are involved in activating RIPK3-dependent necroptosis in the absence of the RIPK1 RHIM domain. Taken together, our results revealed an important role of the RIPK1 RHIM domain in counteracting ZBP1-mediated activation of RIPK3–MLKL-dependent necroptosis, which is critical for preventing lethality during late embryogenesis and skin inflammation in adult mice. These findings identify ZBP1 as a potent inducer of inflammation beyond its role in anti-viral defence^{24,29} and suggest that it could be implicated in inflammatory diseases. Future studies will be required to elucidate the mechanism of ZBP1 activation and how RIPK1 inhibits it, but also its potential implication in the pathogenesis of human diseases.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 June; accepted 21 October 2016.

Published online 7 November 2016.

- Christofferson, D. E., Li, Y. & Yuan, J. Control of life-or-death decisions by RIP1 kinase. *Annu. Rev. Physiol.* **76**, 129–150 (2014).
- Newton, K. RIPK1 and RIPK3: critical regulators of inflammation and cell death. *Trends Cell Biol.* **25**, 347–353 (2015).
- Pasparakis, M. & Vandenabeele, P. Necroptosis and its role in inflammation. *Nature* **517**, 311–320 (2015).
- Silke, J., Rickard, J. A. & Gerlic, M. The diverse role of RIP kinases in necroptosis and inflammation. *Nat. Immunol.* **16**, 689–697 (2015).
- Weinlich, R. & Green, D. R. The two faces of receptor interacting protein kinase-1. *Mol. Cell* **56**, 469–480 (2014).
- Chan, F. K., Luz, N. F. & Moriwaki, K. Programmed necrosis in the cross talk of cell death and inflammation. *Annu. Rev. Immunol.* **33**, 79–106 (2015).
- Lukens, J. R. et al. RIP1-driven autoinflammation targets IL-1 α independently of inflammasomes and RIP3. *Nature* **498**, 224–227 (2013).
- Degterev, A. et al. Chemical inhibitor of nonapoptotic cell death with therapeutic potential for ischemic brain injury. *Nat. Chem. Biol.* **1**, 112–119 (2005).
- He, S. et al. Receptor interacting protein kinase-3 determines cellular necrotic response to TNF- α . *Cell* **137**, 1100–1111 (2009).
- Wang, L., Du, F. & Wang, X. TNF- α induces two distinct caspase-8 activation pathways. *Cell* **133**, 693–703 (2008).
- Berger, S. B. et al. Cutting Edge: RIP1 kinase activity is dispensable for normal development but is a key regulator of inflammation in SHARPIN-deficient mice. *J. Immunol.* **192**, 5476–5480 (2014).
- Polykratis, A. et al. Cutting edge: RIPK1 Kinase inactive mice are viable and protected from TNF-induced necroptosis in vivo. *J. Immunol.* **193**, 1539–1543 (2014).
- Cho, Y. S. et al. Phosphorylation-driven assembly of the RIP1–RIP3 complex regulates programmed necrosis and virus-induced inflammation. *Cell* **137**, 1112–1123 (2009).
- Dannappel, M. et al. RIPK1 maintains epithelial homeostasis by inhibiting apoptosis and necroptosis. *Nature* **513**, 90–94 (2014).
- Dillon, C. P. et al. RIPK1 blocks early postnatal lethality mediated by caspase-8 and RIPK3. *Cell* **157**, 1189–1202 (2014).
- Kaiser, W. J. et al. RIP1 suppresses innate immune necrosis as well as apoptotic cell death during mammalian parturition. *Proc. Natl Acad. Sci. USA* **111**, 7753–7758 (2014).
- Rickard, J. A. et al. RIPK1 regulates RIPK3–MLKL-driven systemic inflammation and emergency hematopoiesis. *Cell* **157**, 1175–1188 (2014).

18. Takahashi, N. *et al.* RIPK1 ensures intestinal homeostasis by protecting the epithelium against apoptosis. *Nature* **513**, 95–99 (2014).
19. Kaiser, W. J., Upton, J. W. & Mocarski, E. S. Receptor-interacting protein homotypic interaction motif-dependent control of NF- κ B activation via the DNA-dependent activator of IFN regulatory factors. *J. Immunol.* **181**, 6427–6434 (2008).
20. Rebsamen, M. *et al.* DAI/ZBP1 recruits RIP1 and RIP3 through RIP homotypic interaction motifs to activate NF- κ B. *EMBO Rep.* **10**, 916–922 (2009).
21. Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501–505 (2007).
22. Schwartz, T., Behlke, J., Lowenhaupt, K., Heinemann, U. & Rich, A. Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat. Struct. Biol.* **8**, 761–765 (2001).
23. Fu, Y. *et al.* Cloning of DLM-1, a novel gene that is up-regulated in activated macrophages, using RNA differential display. *Gene* **240**, 157–163 (1999).
24. Upton, J. W., Kaiser, W. J. & Mocarski, E. S. DAI/ZBP1/DLM-1 complexes with RIP3 to mediate virus-induced programmed necrosis that is targeted by murine cytomegalovirus vIRA. *Cell Host Microbe* **11**, 290–297 (2012).
25. Ishii, K. J. *et al.* TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines. *Nature* **451**, 725–729 (2008).
26. Bonnet, M. C. *et al.* The adaptor protein FADD protects epidermal keratinocytes from necroptosis *in vivo* and prevents skin inflammation. *Immunity* **35**, 572–582 (2011).
27. Kelliher, M. A. *et al.* The death domain kinase RIP mediates the TNF-induced NF- κ B signal. *Immunity* **8**, 297–303 (1998).
28. Meng, L., Jin, W. & Wang, X. RIP3-mediated necrotic cell death accelerates systematic inflammation and mortality. *Proc. Natl Acad. Sci. USA* **112**, 11007–11012 (2015).
29. Kuriakose, T. *et al.* ZBP1/DAI is an innate sensor of influenza virus triggering the NLRP3 inflammasome and programmed cell death pathways. *Sci. Immunol.* **1**, aag2045–aag2045 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Genentech and V. Dixit for *Ripk3*^{−/−} mice, S. Akira for *Zbp1*^{−/−} mice, and J. Han for lentiviral vectors. We thank B. Zevnik, P. Jankowski and S. Assenmacher at the CECAD Transgenic Core Facility for CRISPR/Cas9 mutagenesis in mouse zygotes and C. Uthoff-Hachenberg, J. Buchholz, E. Mahlberg and B. Kühnel for excellent technical assistance. Research reported in this publication was supported by funding from the ERC (grant agreement no. 323040) and the DFG (SFB829 and SFB670). J.L. was supported by a Humboldt research fellowship and C.K. was supported by a Humboldt research fellowship and an EMBO long-term fellowship.

Author Contributions J.L. designed and generated the *Ripk1*^{mRHIM} mice. J.L. and C.K. analysed *Ripk1*^{mRHIM/mRHIM} and *RIPK1*^{mRHIM/E-KO} mice and performed genetic crosses to address the role of RIPK3 and ZBP1 in these mice. J.L. and C.K. carried out all immunoblots and immunoprecipitation experiments. S.K. generated and characterized *RIPK1*^{E-KO} *Zbp1*^{−/−} and *FADD*^{E-KO} *Zbp1*^{−/−} mice and made the initial discovery that ZBP1 is required for keratinocyte necroptosis in *RIPK1*^{E-KO} mice. S.K. and T.-M.V. conducted immunostainings and qRT-PCR assays in skin samples from *RIPK1*^{E-KO} and *RIPK1*^{mRHIM/E-KO} mice. A.P. generated the *Ripk1*^{fl/fl} mice and L.W. generated *Mkl1*^{−/−} mice. M.P. supervised the study, interpreted data and wrote the manuscript together with J.L., C.K. and S.K. J.L., C.K. and S.K. contributed equally and their order of appearance in the author list is random.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.P. (pasparakis@uni-koeln.de).

Reviewer Information *Nature* thanks H. Walczak and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Mice. *Ripk1^{fl/fl}* (ref. 14) and *Fadd^{fl/fl}* (ref. 26), *K14-cre³⁰*, *Ripk3^{-/-}* (ref. 31), and *Zbp1^{-/-}* (ref. 25) mice were described previously. Mice were maintained at the SPF animal facilities of the Institute for Genetics and the CECAD Research Center of the University of Cologne, under a 12 h light cycle, and given a regular chow diet (Harlan, diet number 2918 or Prolab Isopro RMH3000 5P76) *ad libitum*. All animal procedures were conducted in accordance with European, national and institutional guidelines and protocols were approved by local government authorities (Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen, Germany). Animals requiring medical attention were provided with appropriate care and were killed when they developed macroscopically visible skin lesions to minimize suffering. No other exclusion criteria existed. Mice of the indicated genotype were assigned at random to groups. Mouse studies as well as immunohistochemical assessment of pathology were performed in a blinded fashion.

Generation of *Ripk1^{mRHM}* and *Mkl1^{-/-}* mice using Crispr/Cas9-mediated gene targeting in mouse zygotes. For the generation of *Ripk1^{mRHM}* mice Cas9 mRNA (TriLink) together with the 129 bp ssDNA repair oligonucleotides (IDT) and the short guide RNA (sgRNA) targeting the RHIM domain of the murine *Ripk1* gene were microinjected into the pronucleus of fertilized oocytes obtained from C57BL/6 mice. For the generation of the *Mkl1^{-/-}* allele Cas9 mRNA together with the sgRNA targeting the *Mkl1* gene were microinjected into the pronucleus of fertilized oocytes obtained from C57BL/6 mice. On the next day, the injected embryos were transferred to foster mothers and allowed to develop to term. Mutations in the genome of progeny were determined by analysis of genomic DNA using the T7 endonuclease I assay (NEB) and sequencing. For the analysis of the *Mkl1* locus an additional ApaI digest was performed. The sequence of the ssDNA oligo used as a repair template for the *Ripk1* RHIM domain is: 5'-TATCTCTTTTCTATTCAGATGACCTCATAAAATATACTATATCAATAGTTCTGGTATTGTCAGCAGCTAACCAATATATGGATGTTGGACTGAATTCACAACCAACAACAATCTTGCAAAGAA-3'. sgRNA was generated by *in vitro* transcription (NEB, E2040S) from the px330 vector (42230, Addgene) containing the *Ripk1* targeting sequence: 5'-aatagttctggtattcagat-3' or the *Mkl1* targeting sequence: 5'-CGTCTAGGAAACCGTGTGCA-3'. An *Mkl1* allele shown to have a 2bp deletion that causes a frameshift and a premature stop was propagated as the *Mkl1* knockout allele used for this study.

Histological analysis of tissue sections. Skin and intestine tissues were embedded in paraffin or snap frozen in OCT compound. Antigen retrieval for paraffin sections was performed in citrate buffer, pH 6 for the skin sections from *RIPK1^{E-KO}* and *RIPK1^{mRHM/E-KO}* mice and in Tris-EDTA buffer, pH 9 or Proteinase K for the skin and intestine sections from *Ripk1^{mRHM}* and *Ripk1^{-/-}* mice. Anti-active caspase 3 (9661, Cell signalling), anti-F4/80 (clone A3-1, homemade or MCA497G, BIO-RAD), anti-keratin 14 (MS-115, Neomarkers), anti-keratin 6 (PRB-169P, Covance), anti-keratin 10 (PRB-159P, Covance) were used for the staining. Staining was visualized with ABC Kit Vectastain Elite (Vector Laboratories) or streptavidin-HRP (Millipore) and DAB substrate (DAKO and Vector Laboratories) or Alexa-488 and Alexa-549 fluorescence conjugated secondary antibody. F4/80 staining was done on cryo sections. All sections were counterstained with haematoxylin or stained with DAPI. TUNEL assay was performed using DeadEnd Fluorometric TUNEL System (G3250, Promega) as per the manufacturer's instructions. Quantification of epidermal thickness was performed by measurement of epidermal thickness in five optical fields per section. In each field, four measurements were performed. Percentage of inflamed area was determined as the percentage of inflamed versus total number of optical fields at 20× on individual skin sections. TUNEL and CC3 positive cells were quantified under 3–4 optical fields and normalized over the visual area counted.

Immunoblotting and immunoprecipitation. Antibodies against the following proteins were used for western blot analysis: RIPK3 (ADI-905-242-100, Enzo or 95702, Cell signaling), RIPK1 (610459, BD or 3493, Cell signaling), Tubulin (T6074, Sigma), p-IkBα (9246, Cell signaling), IkBα (sc-371, SantaCruz), p-p65 (3033, Cell signaling), Flag (F7425, Sigma), ZBP1 (AG-20B-0010, Adipogen or custom-made, Eurogentec), p-RIPK3 (ab195117, Abcam), Caspase 8 (4790, Cell signaling), MLKL (MABC604, Millipore), GAPDH (NB300-221, Novus). The signals were detected by SuperSignal West Pico Chemiluminescent substrate

(34080, Thermo). The membranes were reprobed after incubation in Restore Western Blot stripping buffer (21059, Thermo).

In order to co-immunoprecipitate ZBP1 interacting proteins, ZBP1 cDNA was generated from IFNβ (97265, Biomol) stimulated wild-type primary MEFs using Superscript III first-strand kit (18080-051, Invitrogen) and the sequence confirmed DNA was cloned into lentiviral pBOB-Flag vector (provided by J. Han, Xiamen University). Primary wild-type and *Ripk1^{mRHM/mRHM}* MEFs were transduced with recombinant lentiviruses. Four days after infection, the cell lysates were prepared in immunoprecipitation buffer (20 mM HEPES-KOH (pH 7.6), 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 10% Glycerol). Flag tag immunoprecipitation was performed by anti-DYKDDDDK antibody magnetic beads (017-25151, Wako) and RIPK1 was precipitated by antibody against RIPK1 (610459, BD) which was BS3-crosslinked to Dynabeads (10004D, Life Technologies).

Cell death assay. Primary MEFs and fetal-liver-derived macrophages were seeded in 96 well plates (1×10^4 cells per well for MEF and 2×10^4 cells per well for FLM) one day before TNF treatment. On the experiment day, indicated amounts of recombinant murine TNF (VIB Protein Service Facility, Ghent), cycloheximide (A0879, Applchem), Z-VAD-FMK (N-1560, Bachem) and Nec-1 (BML-AP309, Enzo) were added to cells. Eighteen hours after incubation, cell viability was determined by neutral red assay as described³².

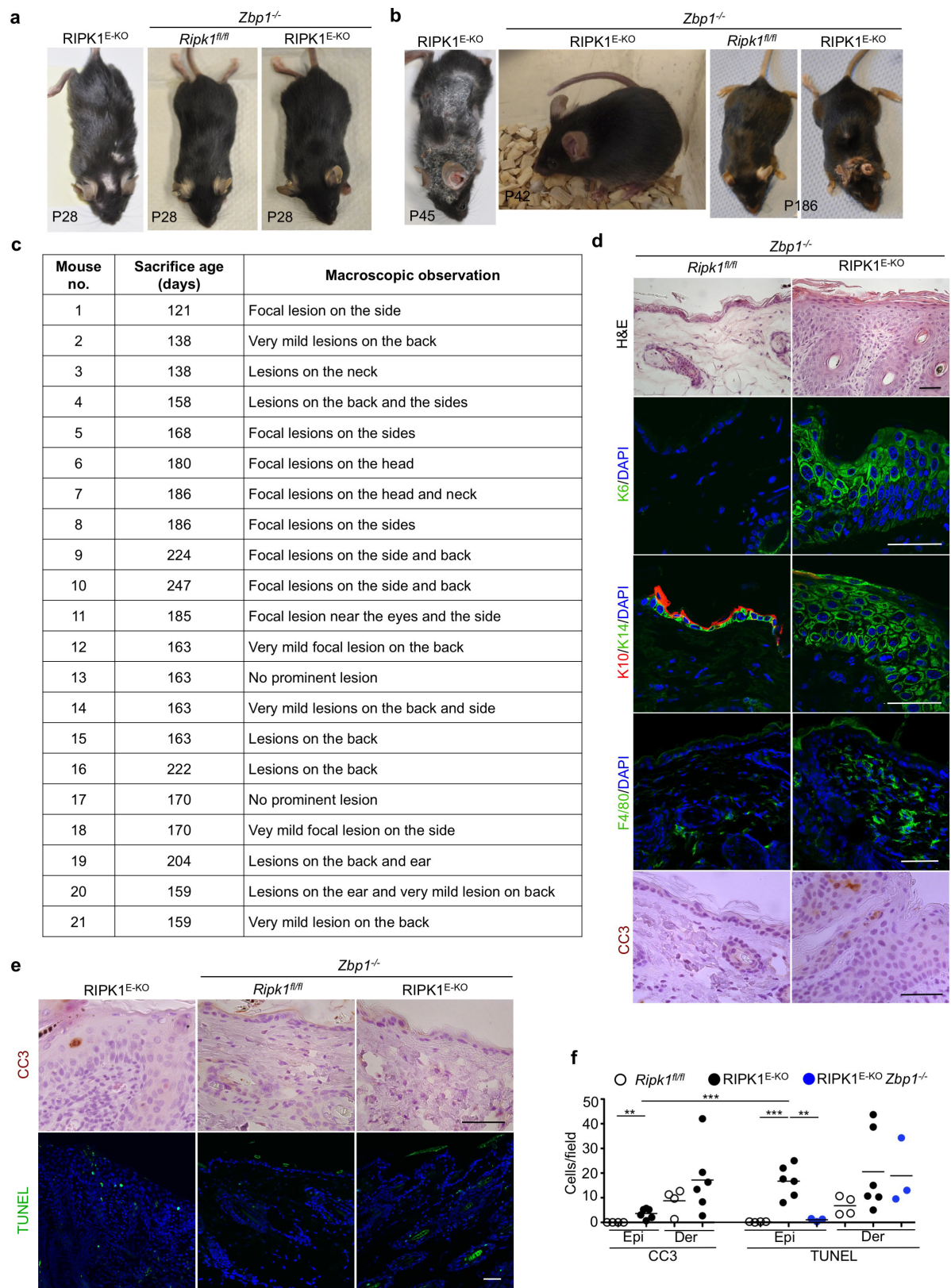
Keratinocytes isolation, culture and stimulation. Keratinocytes from newborn pups were isolated using dispase II (D4693, Sigma). The skin was incubated in dispase II overnight at 4°C. After incubation, epidermis was separated and incubated with TrypLE (12605-010, Gibco) for 20 min and flushed with medium, centrifuged and cultured in low Ca²⁺ DMEM/Ham's F12 medium (F 9092-0.46, Biochrom) with 10% chelax treated FCS and supplements. For immunoblot analyses, 4×10^5 cells were seeded in collagen coated 6 well plates and three hours before stimulation the medium was replaced by fresh medium without EGF. Keratinocytes were stimulated by 20 ng ml⁻¹ TNF or 1,000 U ml⁻¹ IFNβ for eighteen hours.

Quantitative RT-PCR. Total RNA from skin tissue was extracted with Trizol Reagent (Life Technologies) and RNeasy Columns (Qiagen) and cDNA was prepared with Superscript III cDNA-synthesis Kit (Life Technologies). qRT-PCR of *Il1b*, *Il6*, *Il33*, *Tnf*, *Cxcl3*, *Ccl3*, *Ccl5* and *Ccl4* genes was performed with TaqMan probes (Life Technologies). The expression of *Zbp1*, *Ifnb1* and *Ppia* in Fig. 4e were analysed using SYBR Green master mix (4367659, Thermo) with the following primer sets: *Zbp1* forward: 5'-GCTATGACGGACAGACGTGG-3', *Zbp1* reverse: 5'-TGTTGACCGGATTGTGCTGA-3', *Ifnb1* forward: 5'-AGCTCCAAGAAAGGACGAACA-3', *Ifnb1* reverse: 5'-GCCCTGTAGGTGAGGTTGATCT-3', *Ppia* forward: 5'-ATGGTCAACCCACCGTGT-3', *Ppia* reverse: 5'-TTTCTGCTGTCTTTGGAACCTTGTG-3'. HPRT and Tbp for TaqMan and PPIA for SYBR Green were used as reference genes. Data were analysed according to the ΔC_t method.

Statistical analysis. Data shown in column graphs represent mean ± s.e.m. To determine group size necessary for adequate statistical power, power analysis was performed using preliminary data sets. For statistical analysis of data from qRT-PCR, quantification of epidermal thickness and inflamed area, when data did not fulfil the criteria for Gaussian distribution, nonparametric Mann-Whitney test was performed. Unpaired Student's *t*-test was performed for the quantification of TUNEL and CC3 staining. Paired Student's *t*-test was performed for statistical analysis of results obtained from cell viability assays. **P* ≤ 0.05; ***P* ≤ 0.01; ****P* ≤ 0.005. Statistical analysis was performed using Graphpad Prism and Microsoft Excel.

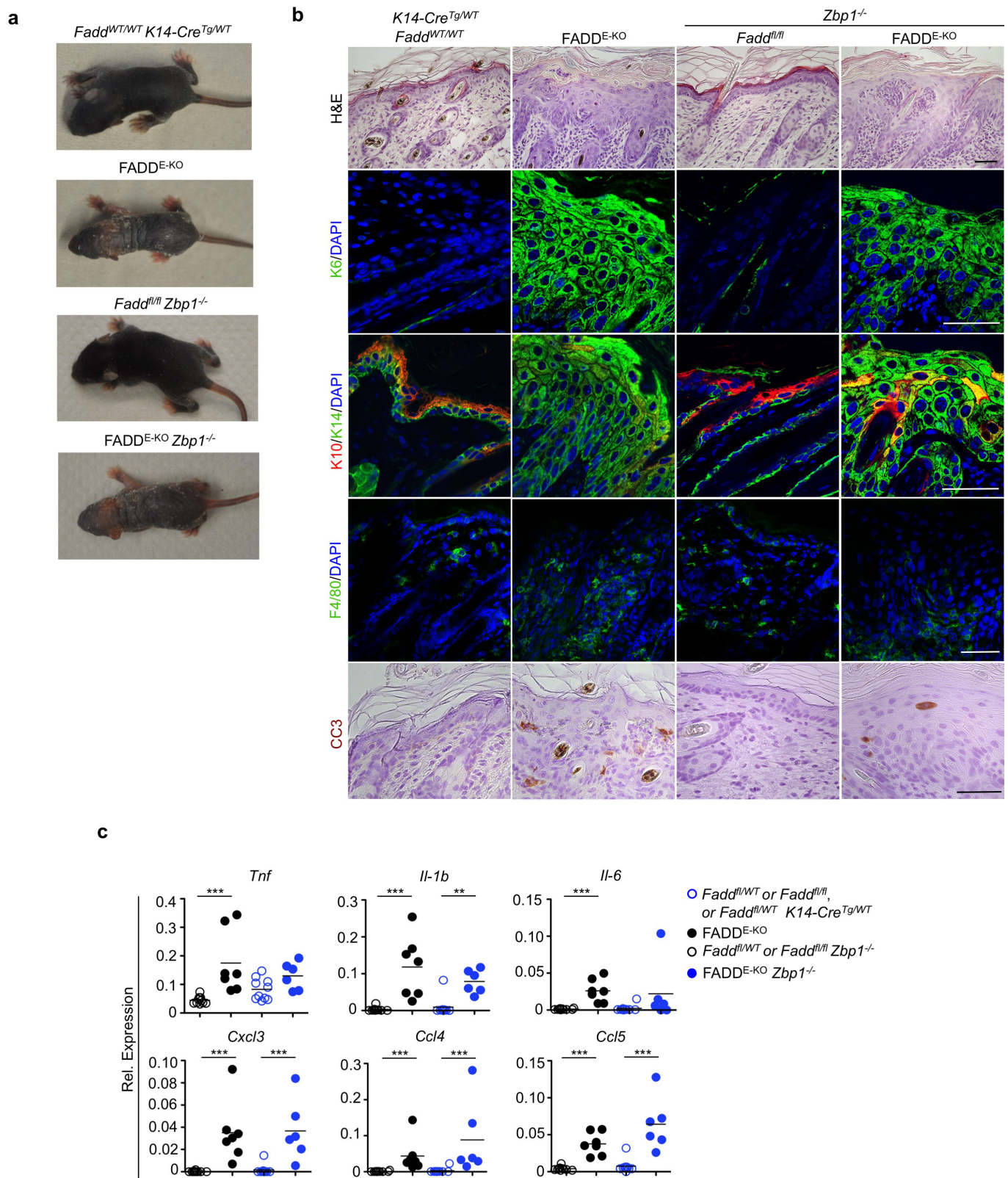
Data availability statement. The authors declare that the data supporting the findings of this study are available within the paper. Source data for all figures are provided with the paper. The *Ripk1^{mRHM}* mice are available from the corresponding author upon request.

30. Hafner, M. *et al.* Keratin 14 Cre transgenic mice authenticate keratin 14 as an oocyte-expressed protein. *Genesis* **38**, 176–181 (2004).
31. Newton, K., Sun, X. & Dixit, V. M. Kinase RIP3 is dispensable for normal NF-κBs, signaling by the B-cell and T-cell receptors, tumor necrosis factor receptor 1, and Toll-like receptors 2 and 4. *Mol. Cell. Biol.* **24**, 1464–1469 (2004).
32. Repetto, G., del Peso, A. & Zurita, J. L. Neutral red uptake assay for the estimation of cell viability/cytotoxicity. *Nat. Protocols* **3**, 1125–1131 (2008).



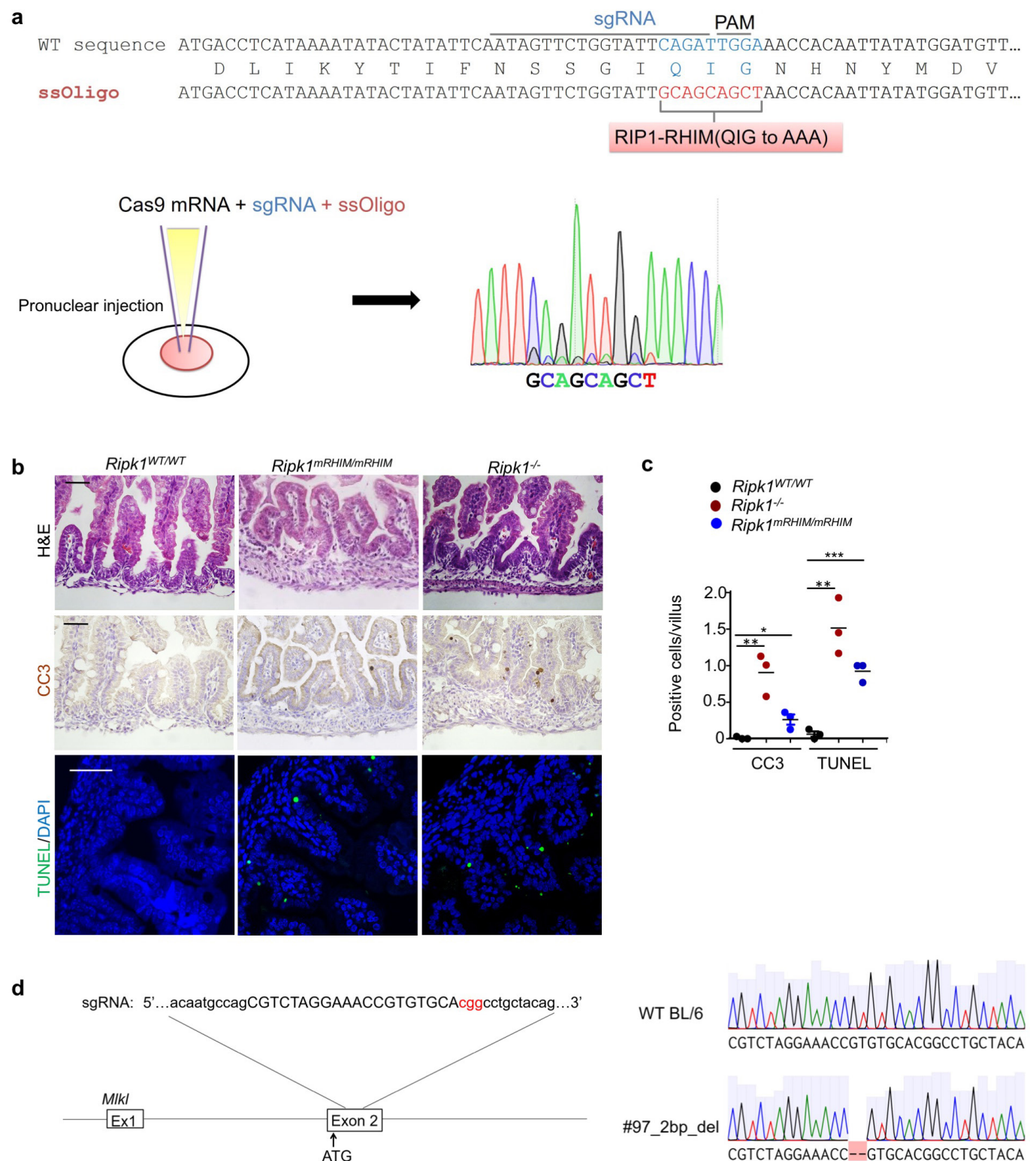
Extended Data Figure 1 | ZBP1 deficiency strongly delays and ameliorates skin inflammation in RIPK1^{E-KO} mice. **a**, Photographs of mice with the indicated genotypes at the age of 4 weeks. Images shown are representative of $n \geq 60$ RIPK1^{E-KO} and $n \geq 40$ RIPK1^{E-KO} *Zbp1*^{-/-} mice. **b**, Photographs of mice with the indicated genotypes and age. Images shown are representative of $n \geq 4$ RIPK1^{E-KO} mice at the age of 5–7 weeks and $n \geq 20$ RIPK1^{E-KO} *Zbp1*^{-/-} mice at the age of 17–35 weeks. **c**, Table summarizing the macroscopically observed skin lesions and time of sacrifice of 21 aged RIPK1^{E-KO} *Zbp1*^{-/-} mice. **d**, Representative images of skin sections from RIPK1^{E-KO} *Zbp1*^{-/-} mice and their respective

controls stained with H&E ($n \geq 18$) or immunostained with the indicated antibodies ($n \geq 4$) at the age of 17–35 weeks. Nuclei stained with DAPI. Scale bars, 50 μ m. **e**, Representative images of skin sections from 4–5 week old RIPK1^{E-KO} ($n \geq 6$) and RIPK1^{E-KO} *Zbp1*^{-/-} ($n \geq 3$) and their respective control mice stained with TUNEL or immunostained with anti-CC3 antibodies. Nuclei stained with DAPI. Scale bars, 50 μ m. **f**, Microscopic quantification of CC3 and TUNEL positive cells on skin sections from 4–5 week old mice with the indicated genotypes. Epi, epidermis; der, dermis. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.005$.



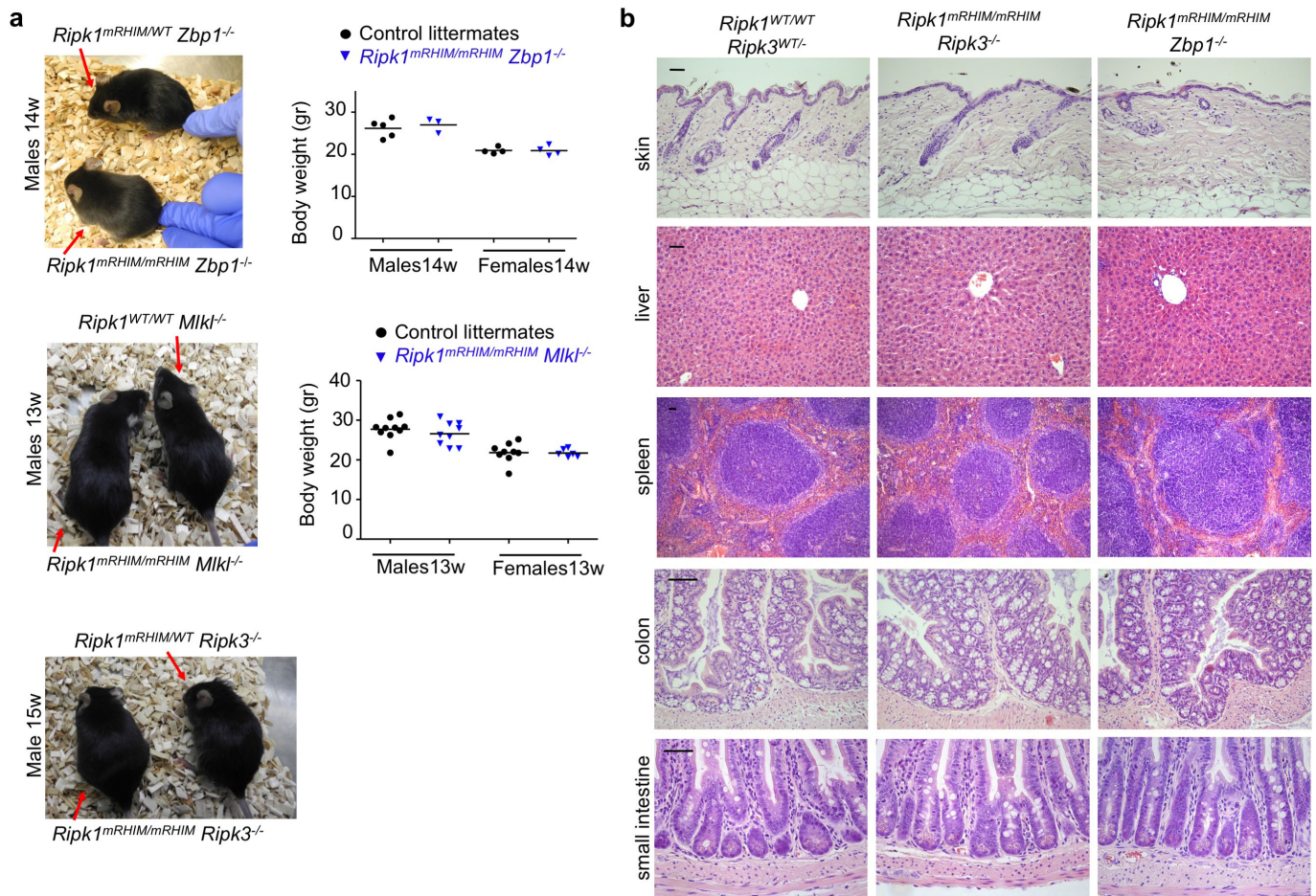
Extended Data Figure 2 | ZBP1 is not required for keratinocyte necroptosis and skin inflammation in mice with epidermis-specific FADD deficiency. **a**, Representative photographs depicting the macroscopically observed phenotype of FADD^{E-KO} ($n \geq 10$) and FADD^{E-KO} *Zbp1*^{-/-} ($n = 5$) mice at the age of 6 days. **b**, Representative images of skin sections from 6 day old mice with the indicated genotypes stained with

H&E ($n = 6$) or immunostained with the indicated antibodies ($n = 3$). Nuclei stained with DAPI. Scale bars, 50 μ m. **c**, qRT-PCR analysis of the mRNA expression of the indicated cytokines and chemokines in total skin RNA from 6-day-old mice with the indicated genotypes. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.005$.

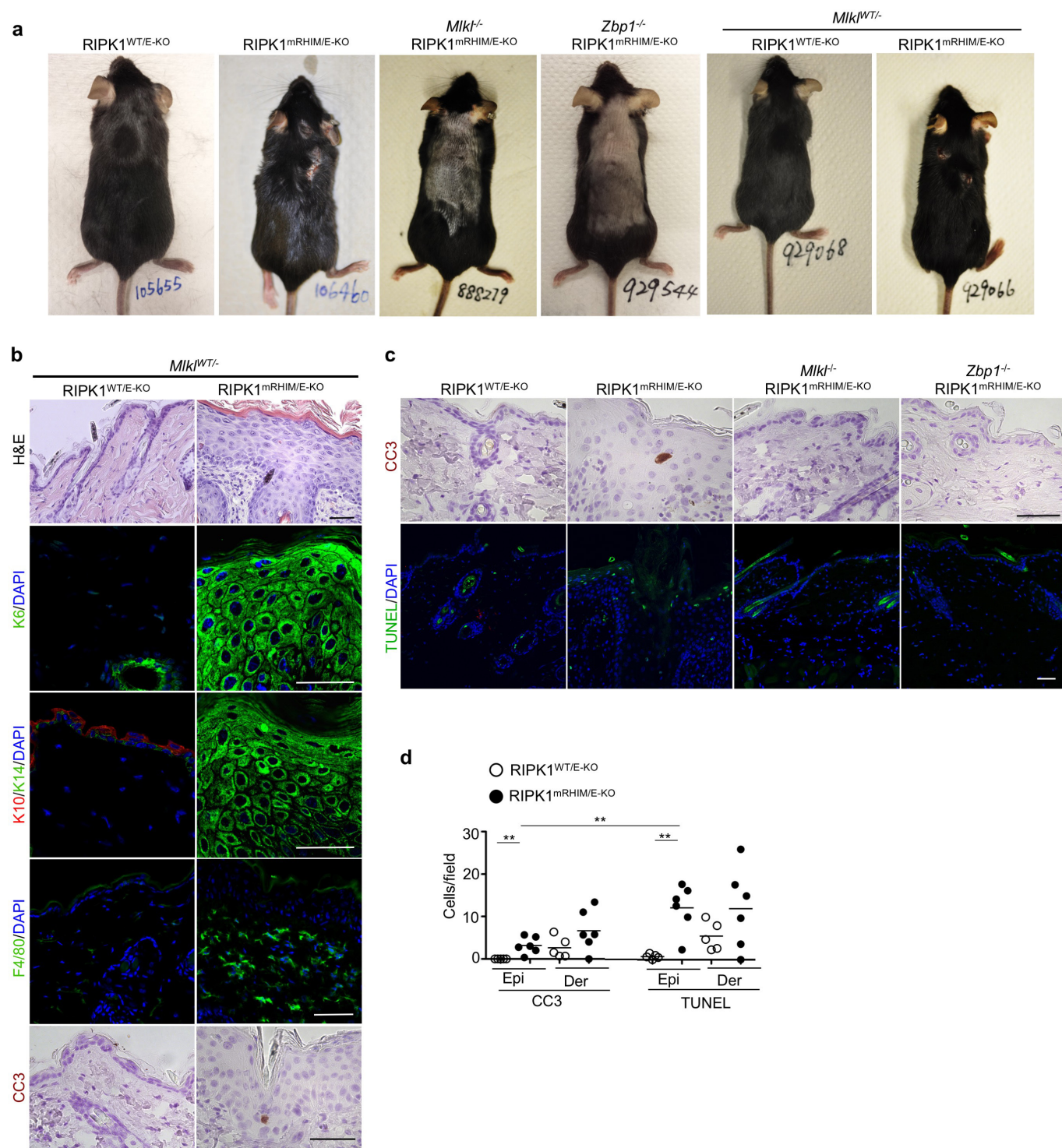


Extended Data Figure 3 | CRISPR-Cas9-mediated generation of *Ripk1*^{mRHIM} and *Mkl*^{-/-} mice. **a**, Schematic depiction of the generation of *Ripk1*^{mRHIM/mRHIM} mice indicating the sequence of the sgRNA and the single stranded oligonucleotide used for mutating the RHIM domain that were introduced by pronuclear injection into mouse zygotes and the sequencing result of one of the two obtained founders. **b**, Small intestinal sections from E18.5 pups were stained with H&E or TUNEL or immunostained with anti-CC3 antibodies. Representative images shown (wild type $n = 6$ for H&E, $n = 3$ for TUNEL and anti-CC3; *Ripk1*^{mRHIM/mRHIM} $n = 5$ for H&E, $n = 3$ for TUNEL and anti-CC3; *Ripk1*^{-/-} $n = 3$ for

H&E, TUNEL and anti-CC3). Nuclei stained with DAPI. Scale bars, 50 μ m. **c**, Microscopic quantification of CC3⁺ and TUNEL⁺ cells on gut sections from E18.5 pups with the indicated genotypes. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.005$. **d**, Diagram indicating the sgRNA target sequence (capital letters) used to generate a mutation in exon 2 downstream of the ATG of the *Mkl* gene. The PAM sequence is indicated in red. Sequencing showing the 2 base pair deletion found in 97 at position chr8:111,333,648–111,333,649 (mm10), which results in a frameshift after amino acid 34 and a premature stop codon at amino acid position 55 of MLKL. This *Mkl* knockout allele was used throughout this study.

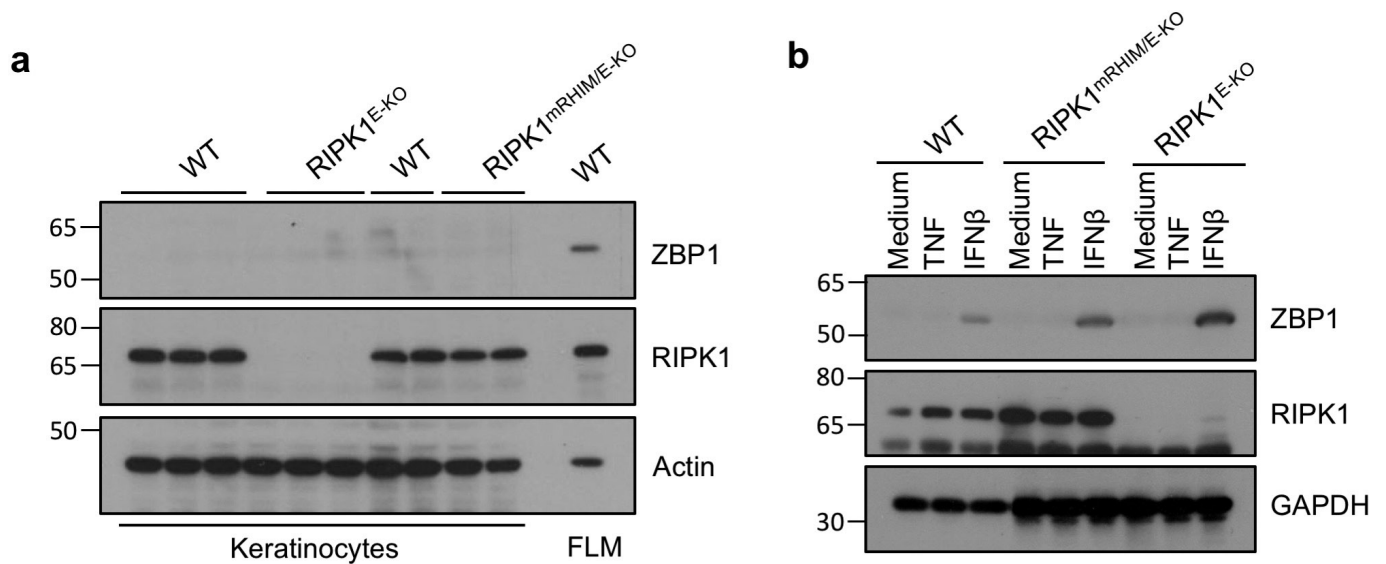


Extended Data Figure 4 | Rescue of perinatal lethality of *Ripk1^{mRHIM/mRHIM}* mice by deficiency of ZBP1, MLKL or RIPK3. a, Representative photographs and body weights of the indicated mice. b, Representative H&E stainings of skin, liver, spleen, colon and small intestine sections from 5-month-old *Ripk3^{WT/-}* ($n = 4$), *Ripk1^{mRHIM/mRHIM} Ripk3^{-/-}* ($n = 4$) and *Ripk1^{mRHIM/mRHIM} Zbp1^{-/-}* mice ($n = 3$).



Extended Data Figure 5 | MLKL or ZBP1-deficiency prevents skin inflammation in $RIPK1^{mRHIM/E-KO}$ mice. **a**, Representative photographs of $RIPK1^{mRHIM/E-KO}$ ($n=9$), $RIPK1^{mRHIM/E-KO} Mikl^{WT/-}$ ($n=11$), $RIPK1^{mRHIM/E-KO} Mikl^{-/-}$ ($n=16$) and $RIPK1^{mRHIM/E-KO} Zbp1^{-/-}$ ($n=7$) at the age of 9–11 weeks. **b**, Representative images of skin sections from 9–11-week-old $RIPK1^{mRHIM/E-KO} Mikl^{WT/-}$ ($n=11$) and the respective control mice stained with H&E or immunostained with the indicated antibodies. Nuclei stained with DAPI. Scale bars, 50 μm . **c**, Representative

images of skin sections from 9–11-week-old $RIPK1^{mRHIM/E-KO}$ ($n \geq 6$), $RIPK1^{mRHIM/E-KO} Mikl^{-/-}$ ($n=3$), $RIPK1^{mRHIM/E-KO} Zbp1^{-/-}$ ($n=3$) and their respective control mice stained with TUNEL or immunostained with anti-CC3 antibodies. Nuclei stained with DAPI. Scale bars, 50 μm . **d**, Microscopic quantification of CC3⁺ and TUNEL⁺ positive cells on skin sections from 9–11 week old mice with the indicated genotypes. Epi, epidermis; der, dermis. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.005$.



Extended Data Figure 6 | Expression of ZBP1 in primary keratinocytes.
a, Immunoblot analysis of lysates of primary keratinocytes derived from mice with the indicated genotypes. Each lane represents keratinocytes from individual mice. Cell lysates of wild-type FLM was used as a positive

control. For gel source data, see Supplementary Fig. 1. **b**, Immunoblot analysis of primary keratinocytes derived from mice with the indicated genotypes were left untreated ('medium') or stimulated with TNF or IFN- β for 18 h. For gel source data, see Supplementary Fig. 1.

Extended Data Table 1 | Numbers of progeny obtained at weaning age from rescue crosses of *Ripk1*^{mRHIM/WT} mice with *Ripk3*^{-/-}, *Mkl1*^{-/-}, *Zbp1*^{-/-} and *Trif*^{-/-} mice

Genotype of offspring Genotype of parents		<i>Ripk3</i> ^{-/-}			<i>Ripk3</i> ^{WT/-}		
		<i>Ripk1</i> ^{WT/WT}	<i>Ripk1</i> ^{mRHIM/WT}	<i>Ripk1</i> ^{mRHIM/mRHIM}	<i>Ripk1</i> ^{WT/WT}	<i>Ripk1</i> ^{mRHIM/WT}	<i>Ripk1</i> ^{mRHIM/mRHIM}
Line 1	<i>Ripk1</i> ^{mRHIM/WT} <i>Ripk3</i> ^{WT/-}	6	12	9	12	27	13
	<i>Ripk1</i> ^{mRHIM/WT} <i>Ripk3</i> ^{WT/-}						
Line 2	<i>Ripk1</i> ^{mRHIM/WT} <i>Ripk3</i> ^{WT/-}	16	32	10	18	33	5
	<i>Ripk1</i> ^{mRHIM/WT} <i>Ripk3</i> ^{-/-}						

Genotype of offspring Genotype of parents		<i>Mkl1</i> ^{-/-}			<i>Mkl1</i> ^{WT/-}		
		<i>Ripk1</i> ^{WT/WT}	<i>Ripk1</i> ^{mRHIM/WT}	<i>Ripk1</i> ^{mRHIM/mRHIM}	<i>Ripk1</i> ^{WT/WT}	<i>Ripk1</i> ^{mRHIM/WT}	<i>Ripk1</i> ^{mRHIM/mRHIM}
Line 2	<i>Ripk1</i> ^{mRHIM/WT} <i>Mkl1</i> ^{WT/-}	17	41	13	10	32	0
	<i>Ripk1</i> ^{mRHIM/WT} <i>Mkl1</i> ^{-/-}						
	<i>Ripk1</i> ^{mRHIM/WT} <i>Mkl1</i> ^{-/-}	16	25	20			
	<i>Ripk1</i> ^{mRHIM/WT} <i>Mkl1</i> ^{-/-}						

Genotype of offspring Genotype of parents		<i>Zbp1</i> ^{-/-}			<i>Zbp1</i> ^{WT/-}		
		<i>Ripk1</i> ^{WT/WT}	<i>Ripk1</i> ^{mRHIM/WT}	<i>Ripk1</i> ^{mRHIM/mRHIM}	<i>Ripk1</i> ^{WT/WT}	<i>Ripk1</i> ^{mRHIM/WT}	<i>Ripk1</i> ^{mRHIM/mRHIM}
Line 2	<i>Ripk1</i> ^{mRHIM/WT} <i>Zbp1</i> ^{WT/-}	4	8	5	10	14	0
	<i>Ripk1</i> ^{mRHIM/WT} <i>Zbp1</i> ^{WT/-}						
	<i>Ripk1</i> ^{mRHIM/WT} <i>Zbp1</i> ^{WT/-}	10	19	7	4	23	0
	<i>Ripk1</i> ^{mRHIM/WT} <i>Zbp1</i> ^{-/-}						

Genotype of offspring Genotype of parents		<i>Trif</i> ^{-/-}			<i>Trif</i> ^{WT/-}		
		<i>Ripk1</i> ^{WT/WT}	<i>Ripk1</i> ^{mRHIM/WT}	<i>Ripk1</i> ^{mRHIM/mRHIM}	<i>Ripk1</i> ^{WT/WT}	<i>Ripk1</i> ^{mRHIM/WT}	<i>Ripk1</i> ^{mRHIM/mRHIM}
Line 2	<i>Ripk1</i> ^{mRHIM/WT} <i>Trif</i> ^{WT/-}	7	12	0	14	27	0
	<i>Ripk1</i> ^{mRHIM/WT} <i>Trif</i> ^{WT/-}						
	<i>Ripk1</i> ^{mRHIM/WT} <i>Trif</i> ^{WT/-}	11	24	0	16	12	0
	<i>Ripk1</i> ^{mRHIM/WT} <i>Trif</i> ^{-/-}						

RIPK1 inhibits ZBP1-driven necroptosis during development

Kim Newton¹, Katherine E. Wickliffe¹, Allie Maltzman¹, Debra L. Dugger¹, Andreas Strasser^{2,3}, Victoria C. Pham⁴, Jennie R. Lill⁴, Merone Roose-Girma⁵, Søren Warming⁵, Margaret Solon⁶, Hai Ngu⁶, Joshua D. Webster⁶ & Vishva M. Dixit¹

Receptor-interacting protein kinase 1 (RIPK1) promotes cell survival—mice lacking RIPK1 die perinatally, exhibiting aberrant caspase-8-dependent apoptosis and mixed lineage kinase-like (MLKL)-dependent necroptosis^{1–3}. However, mice expressing catalytically inactive RIPK1 are viable^{2,4,5}, and an ill-defined pro-survival function for the RIPK1 scaffold has therefore been proposed. Here we show that the RIP homotypic interaction motif (RHIM) in RIPK1 prevents the RHIM-containing adaptor protein ZBP1 (Z-DNA binding protein 1; also known as DAI or DLM1) from activating RIPK3 upstream of MLKL. *Ripk1*^{RHIM/RHIM} mice that expressed mutant RIPK1 with critical RHIM residues IQIG mutated to AAAA died around birth and exhibited RIPK3 autophosphorylation on Thr231 and Ser232, which is a hallmark of necroptosis⁶, in the skin and thymus. Blocking necroptosis with catalytically inactive RIPK3(D161N), RHIM mutant RIPK3, RIPK3 deficiency, or MLKL deficiency prevented lethality in *Ripk1*^{RHIM/RHIM} mice. Loss of ZBP1, which engages RIPK3 in response to certain viruses^{7,8} but previously had no defined role in development, also prevented perinatal lethality in *Ripk1*^{RHIM/RHIM} mice. Consistent with the RHIM of RIPK1 functioning as a brake that prevents ZBP1 from engaging the RIPK3 RHIM, ZBP1 interacted with RIPK3 in *Ripk1*^{RHIM/RHIM} *Mkl1*^{−/−} macrophages, but not in wild-type, *Mkl1*^{−/−} or *Ripk1*^{RHIM/RHIM} *Ripk3*^{RHIM/RHIM} macrophages. Collectively, these findings indicate that the RHIM of RIPK1 is critical for preventing ZBP1/RIPK3/MLKL-dependent necroptosis during development.

Studies of RIPK1 deficiency show that RIPK1 suppresses RIPK3/MLKL-dependent necroptosis in some cell types, and FADD (FAS-associated via death domain)/caspase-8-dependent apoptosis in others^{1–3,9–12}. However, mutation of critical catalytic residues in RIPK1 or inhibition of RIPK1 enzymatic activity does not promote cell death^{2,4,5,11,12}. Therefore, RIPK1 is hypothesized to form a pro-survival scaffold that is independent of its kinase activity. In addition to its kinase domain, RIPK1 has RHIM and death domain (DD) protein interaction motifs¹³ (Fig. 1a). The DD mediates recruitment of RIPK1 to proteins such as tumour necrosis factor receptor 1 (TNFR1) and FADD, whereas the RHIM mediates interactions with RIPK3 (ref. 13). To determine if the RHIM of RIPK1 is an essential part of the RIPK1 pro-survival function, we generated *Ripk1*^{RHIM/RHIM} mice, in which RHIM core residues IQIG¹⁴ are mutated to AAAA (Fig. 1a).

Ripk1^{RHIM/RHIM} mice died around birth (Fig. 1b). Abnormalities at embryonic day 18.5 (E18.5) included epidermal hyperplasia, dermatitis and oedema of the skin, increased number of neutrophils in the peritoneum, and decreased hepatocellular vacuolation (Fig. 1c, Extended Data Fig. 1a). Similar skin abnormalities in *Ripk1*^{−/−} embryos require RIPK3- and MLKL-dependent necroptosis¹. Of note, *Ripk1*^{RHIM/RHIM} embryos lacked the caspase-8-dependent intestinal lesions seen in *Ripk1*^{−/−} embryos¹ (Extended Data Fig. 1a). These data

suggest that mutation of the RIPK1 RHIM primarily unleashes the RIPK3–MLKL necroptosis pathway (Extended Data Fig. 1b).

RIPK3 autophosphorylation on Thr231 and Ser232 is a hallmark of necroptosis⁶. A phospho-specific RIPK3 antibody recognizing these sites (Extended Data Fig. 1c, d) stained many cells in the skin and thymus of *Ripk1*^{RHIM/RHIM} embryos, but only rare cells throughout multiple tissues including the bladder, liver, heart and intestine were positive (Fig. 1c, Extended Data Fig. 1e). Importantly, phospho-RIPK3 staining was absent in wild-type embryos. RIPK3 autophosphorylation in *Ripk1*^{RHIM/RHIM} skin coincided with increased cell death on the basis of TUNEL staining, which marks dying cells with DNA damage, and cleaved caspase-3 staining, which marks apoptotic cells (Extended Data Fig. 2a, b). Increased cell death was less pronounced in the *Ripk1*^{RHIM/RHIM} thymus (Extended Data Fig. 2b).

Perinatal lethality and the histologic lesions in *Ripk1*^{RHIM/RHIM} mice were prevented by either RIPK3 or MLKL deficiency, suggesting these defects stemmed from aberrant necroptosis (Fig. 1c–f, Extended Data Figs 1a, 2). Although RIPK3 has been linked to apoptosis induction in some contexts¹³, MLKL has not. Therefore, increased apoptosis in the *Ripk1*^{RHIM/RHIM} dermis may have been a secondary consequence and inflammation driven. *Ripk1*^{RHIM/RHIM} *Ripk3*^{+/-} mice were also viable and fertile (Fig. 1f), although 4 out of 7 mice developed severe dermatitis around the neck and ears after 6–9 months (Extended Data Fig. 2f). Sensitivity to *Ripk3* gene dosage was not unexpected because halving the *Ripk3* gene dosage also averts perinatal lethality in *Ripk1*^{−/−} *Casp8*^{−/−} mice⁵. *Ripk1*^{RHIM/RHIM} mice expressing RIPK3 with RHIM residues VQIG mutated to AAAA (Extended Data Fig. 3a) or catalytically inactive RIPK3(D161N)⁴ also survived beyond weaning, were fertile, and had no overt defects (albeit the eldest mice analysed were only aged 5–8 months) (Extended Data Fig. 3b–d). These data provide further support for necroptosis driving the lesions in *Ripk1*^{RHIM/RHIM} mice because both the RHIM and kinase activity of RIPK3 are critical for necroptosis induction^{15–17} (Extended Data Fig. 3e–h).

RIPK1 deficiency compromises TNF signalling in some cell types^{9,18}, so we compared wild-type, *Ripk1*^{−/−}, and *Ripk1*^{RHIM/RHIM} primary mouse embryo fibroblasts (MEFs) after TNF treatment. Phosphorylation and degradation of I κ B α , and phosphorylation of p38 and JNK were impaired in *Ripk1*^{−/−} MEFs as reported⁸, but occurred normally in *Ripk1*^{RHIM/RHIM} MEFs (Extended Data Fig. 4a, b). *Ripk1*^{RHIM/RHIM} MEFs treated with TNF also maintained normal levels of FLIP (FLICE-inhibitory protein), TRAF2 (TNF receptor associated factor 2), and cIAP1 (cellular inhibitor of apoptosis protein 1), whereas these proteins were lost from *Ripk1*^{−/−} MEFs (Extended Data Fig. 4c). Therefore, these RIPK1-dependent signalling events triggered by TNF in MEFs do not require the RIPK1 RHIM.

Next we examined necroptosis signalling in primary *Ripk1*^{RHIM/RHIM} MEFs treated with TNF plus the pan-caspase inhibitor Z-VAD-FMK

¹Department of Physiological Chemistry, Genentech, 1 DNA Way, South San Francisco, California 94080, USA. ²Molecular Genetics of Cancer Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia. ³Department of Medical Biology, Melbourne University, Victoria 3010, Australia. ⁴Department of Proteomics and Biological Resources, Genentech, 1 DNA Way, South San Francisco, California 94080, USA. ⁵Department of Molecular Biology, Genentech, 1 DNA Way, South San Francisco, California 94080, USA. ⁶Department of Pathology, Genentech, 1 DNA Way, South San Francisco, California 94080, USA.

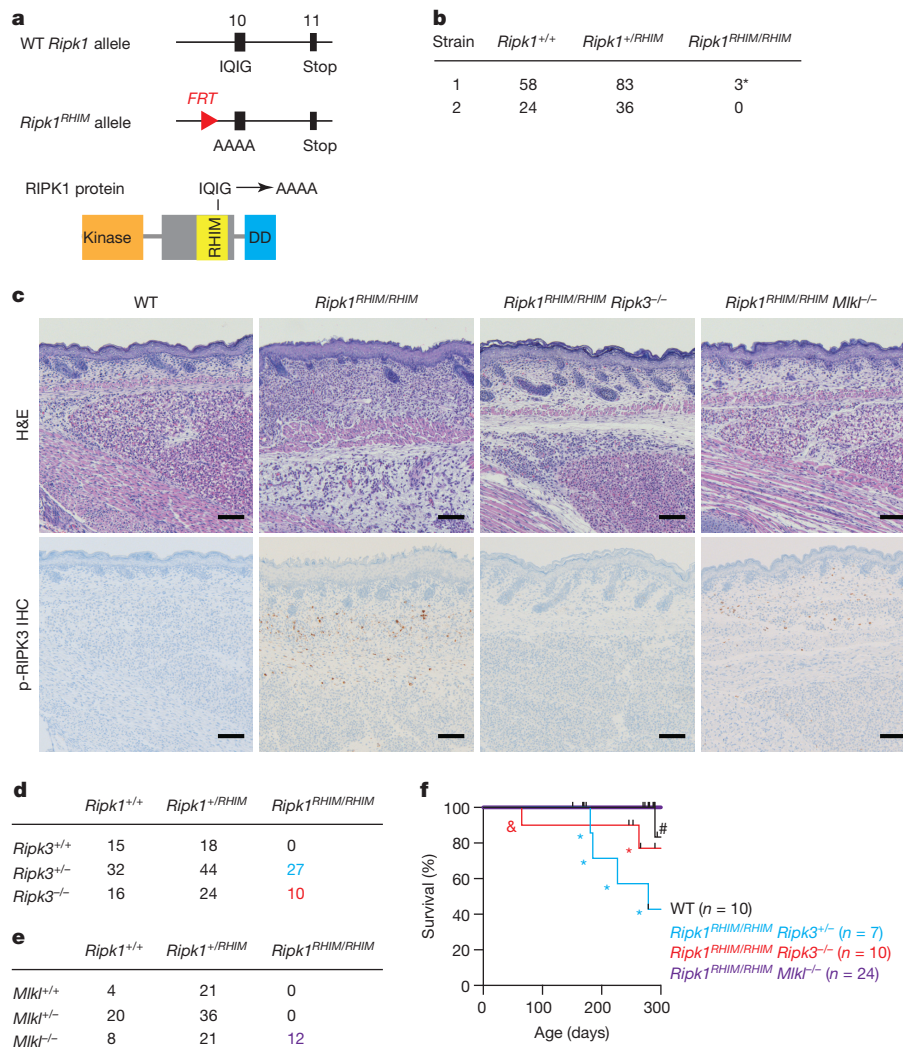


Figure 1 | Mutation of the RIPK1 RHIM causes RIPK3/MLKL-dependent lethality. **a**, Organization of the *Ripk1*^{RHIM} mutant allele. Black boxes indicate exons. **b**, Numbers of offspring from intercrossing *Ripk1*^{+/^{RHIM}} parents. **c**, E18.5 embryo sections. Autophosphorylated RIPK3 stained brown (p-RIPK3 immunohistochemistry (IHC)). H&E, haematoxylin and eosin. Scale bars, 100 μ m. Results are representative

(hereafter denoted as TZ). Similar to wild-type macrophages (Extended Data Figs 3g and 4d), wild-type or *Ripk1*^{RHIM/RHIM} MEFs exhibited RIPK1 phosphorylation on Ser166 after TZ treatment (Fig. 2a). We propose that this modification represents RIPK1 autophosphorylation because it was absent when cells expressed catalytically inactive RIPK1 (D138N). Interestingly, Ser166 phosphorylation was detected as early as 5 min after treatment with TNF or TZ (Fig. 2b), indicating that it is not a necroptosis-specific modification. Analysis of the TNFR1-associated signalling complex induced by TZ (Fig. 2c) indicated that wild-type, catalytically inactive, or RHIM mutant RIPK1 were incorporated into the complex and modified in a manner consistent with them being ubiquitylated¹⁹. However, only wild-type and RHIM mutant RIPK1 were also phosphorylated on Ser166. Ubiquitylation of RIPK1 is thought to contribute to TAK1 and IKK activation^{20–24}, so our data implying that RIPK1 autophosphorylation is dispensable for RIPK1 ubiquitylation fit with TNF activating IKK- and TAK1-dependent signalling pathways normally in *Ripk1*^{D138N/D138N} cells⁴ (Extended Data Fig. 5b).

We explored whether TNF-induced phosphorylation of RIPK1 Ser166 required RIPK1 ubiquitylation by pretreating wild-type MEFs with the IAP (inhibitor of apoptosis protein) antagonist BV6 or by analysing *Tradd*^{-/-} macrophages (Extended Data Fig. 5a, b). Wild-type

of three (wild type, *Ripk1*^{RHIM/RHIM} and *Ripk1*^{RHIM/RHIM} *Ripk3*^{-/-}) or five embryos (*Ripk1*^{RHIM/RHIM} *Mlkl*^{-/-}). **d**, **e**, Numbers of offspring from intercrossing compound heterozygote parents. **f**, Kaplan–Meier plot of mouse survival. Mice were found dead (&) or euthanized because of an enlarged abdomen (#) or skin lesions (*).

MEFs treated with BV6 lacked detectable cIAP1, consistent with BV6 inducing proteasomal degradation of cIAP1 and cIAP2 (ref. 25). In addition, RIPK1 associated with TNFR1 no longer migrated as a high molecular weight smear, consistent with it being poorly ubiquitylated (Extended Data Fig. 5a). However, RIPK1 phosphorylation on Ser166 was still detected, implying that RIPK1 autophosphorylation does not require RIPK1 ubiquitylation. Similar results were obtained using *Tradd*^{-/-} macrophages, which had reduced RIPK1 ubiquitylation in the TNFR1 signalling complex (Extended Data Fig. 5b). Collectively, our data suggest that RIPK1 associated with TNFR1 undergoes ubiquitylation and autophosphorylation with each modification not requiring the other.

TZ caused RIPK1 to interact with RIPK3 in wild-type MEFs, whereas TNF or Z-VAD-FMK alone did not (Extended Data Fig. 5c). Consistent with previous studies^{15,16}, this interaction required both the RHIM and kinase activity of RIPK1 (Fig. 2a). RIPK1 interacting with RIPK3 promotes RIPK3 oligomerization and autophosphorylation²⁶. Therefore, as expected, TZ induced RIPK3 autophosphorylation in wild-type MEFs, but not in *Ripk1*^{D138N/D138N} MEFs (Fig. 2a). RIPK3 was not autophosphorylated in wild-type MEFs treated with TNF or Z-VAD-FMK individually (Fig. 2d). Notably, both *Ripk1*^{RHIM/RHIM} and *Ripk1*^{-/-} MEFs exhibited RIPK3 autophosphorylation in medium alone (Fig. 2a,

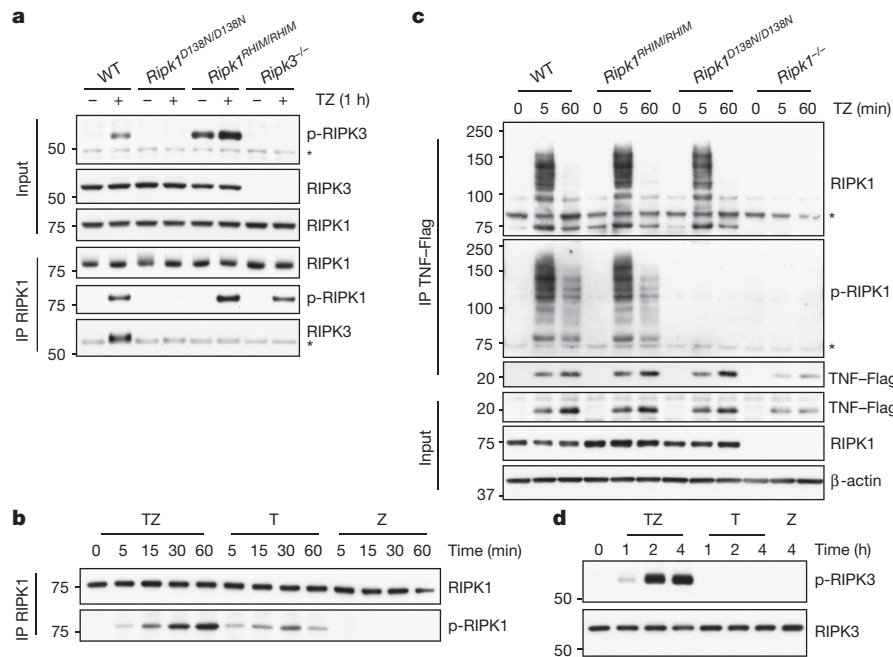


Figure 2 | The RIPK1 RHIM is dispensable for RIPK3 autophosphorylation but mediates interactions with RIPK3 in response to TZ. **a–d**, Western blots of MEFs or immunoprecipitations (IP). T, TNF; Z, Z-VAD-FMK; p-RIPK1, phosphorylated RIPK1 Ser166;

p-RIPK3, phosphorylated RIPK3 Thr231 and Ser232. Asterisks indicate non-specific bands. Results are representative of 2 (**c**), 4 (**a**, **b**) or 5 (**d**) independent experiments. For gel source data, see Supplementary Fig. 1.

Extended Data Fig. 5d). These data indicate that the RHIM of RIPK1 suppresses RIPK3 activation in MEFs as well as in the skin and thymus of the developing mouse. RIPK3 autophosphorylation in the *Ripk1^{RHIM/RHIM}* or *Ripk1^{-/-}* MEFs induced little cell death (Extended Data Fig. 5e), although it should be noted that wild-type MEFs were considerably less sensitive to TZ than either macrophages (Extended Data Fig. 3e) or immortalised MEFs⁴. Primary MEFs expressed less MLKL than macrophages (Extended Data Fig. 5f), which perhaps explains their relative insensitivity to RIPK3 activation.

We next investigated whether an upstream stimulus drives RIPK3 activation in *Ripk1^{-/-}* or *Ripk1^{RHIM/RHIM}* cells. ZBP1 and TRIF (also known as Ticam1) are RHIM-containing proteins that interact with RIPK3, and we therefore tested whether loss of both ZBP1 and TRIF substituted for RIPK3 deficiency and rescued lethality in *Ripk1^{-/-}Casp8^{-/-}* mice^{1–3}. After intercrossing *Ripk1^{+/-}Casp8^{+/-}Zbp1^{-/-}Trif^{-/-}* mice, we obtained viable *Ripk1^{-/-}Casp8^{-/-}Zbp1^{-/-}Trif^{-/-}* mice with a median survival of 17 weeks, whereas *Ripk1^{-/-}Casp8^{+/-}Zbp1^{-/-}Trif^{-/-}* or *Ripk1^{-/-}Casp8^{+/-}Zbp1^{-/-}Trif^{-/-}* littermates died within 3 weeks (Fig. 3a, b). With the exception of three mice with malocclusion, all weaned *Ripk1^{-/-}Casp8^{-/-}Zbp1^{-/-}Trif^{-/-}* mice had to be euthanized after 125 to 175 days when masses caused by lymphadenopathy were observed (Fig. 3c). CD3⁺B220⁺ T cells had accumulated (Fig. 3d) owing to caspase-8 deficiency²⁷. Lymphadenopathy was accompanied by elevated serum cytokines and chemokines including IL-17A and CCL4 (Extended Data Fig. 6a). The only other consistent phenotype revealed by histological analysis of all major organ systems at 3–6 months of age was a marked depletion of testicular germ cells in the *Ripk1^{-/-}Casp8^{-/-}Zbp1^{-/-}Trif^{-/-}* males (Extended Data Fig. 6b).

Survival of *Ripk1^{-/-}Casp8^{-/-}Zbp1^{-/-}Trif^{-/-}* mice beyond weaning suggests that ZBP1 and/or TRIF promote RIPK3 activation when RIPK1 is absent. Consistent with this notion, E18.5 *Ripk1^{-/-}Zbp1^{-/-}Trif^{-/-}* skin had few cells containing autophosphorylated RIPK3 when compared to *Ripk1^{-/-}* skin (Fig. 3e). Epidermal hyperplasia and dermatitis in *Ripk1^{-/-}Zbp1^{-/-}Trif^{-/-}* skin was presumably due to caspase-8-dependent apoptosis because neither occurred in the skin of *Ripk1^{-/-}Casp8^{-/-}Zbp1^{-/-}Trif^{-/-}* mice aged 3–6 months (Extended Data Fig. 6c). One *Ripk1^{-/-}Casp8^{-/-}Zbp1^{-/-}* mouse and one *Ripk1^{-/-}Casp8^{-/-}Zbp1^{-/-}Trif^{-/-}* mouse were obtained

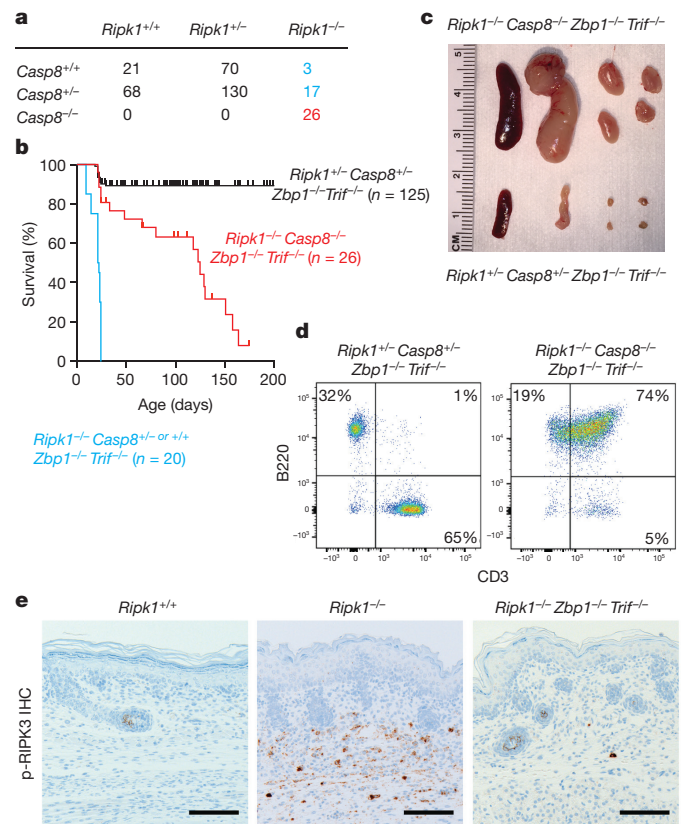


Figure 3 | Combined ZBP1 and TRIF deficiency rescues perinatal lethality in *Ripk1^{-/-}Casp8^{-/-}* mice. **a**, Numbers of offspring from intercrossing *Ripk1^{+/-}Casp8^{+/-}Zbp1^{-/-}Trif^{-/-}* parents. **b**, Kaplan–Meier plot of mouse survival. Animals not found at weaning were recorded as dead at 21 days, but may have died earlier. **c**, Spleen and lymph nodes of male littermates aged 18 weeks. **d**, Flow cytometric analysis of lymph node cells from female mice aged 15–17 weeks. **e**, E18.5 skin sections. Autophosphorylated RIPK3 (p-RIPK3 IHC) stained brown. Scale bar, 100 μ m. Results are representative of 3 (**e**) or 4 (**c**, **d**) mice of each genotype.

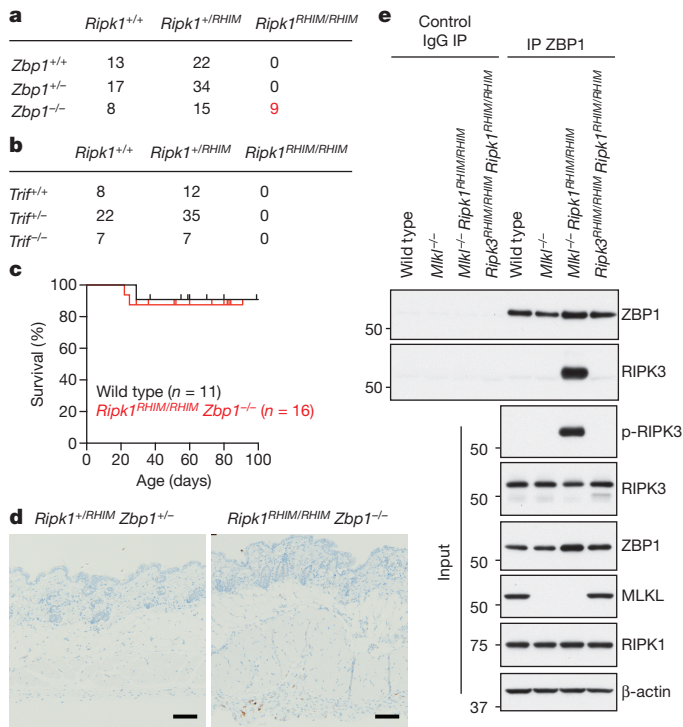


Figure 4 | ZBP1 interacts with RIPK3 to trigger necroptosis in *Ripk1*^{RHIM/RHIM} mice. **a**, **b**, Numbers of offspring from intercrossing compound heterozygote parents. **c**, Kaplan-Meier plot of mouse survival. **d**, Skin sections from female littermates aged 7 weeks. Autophosphorylated RIPK3 (p-RIPK3 IHC) stained brown. Scale bar, 100 μm. Results are representative of 2 *Ripk1*^{RHIM/RHIM} *Zbp1*^{-/-} mice. **e**, Western blots of macrophages or immunoprecipitates (IP). Results are representative of 4 independent experiments. For gel source data, see Supplementary Fig. 1.

from limited intercrossing of *Ripk1*^{+/+} *Casp8*^{+/+} *Zbp1*^{+/+} *Trif*^{+/+} mice. At 7 weeks of age, they both exhibited suppurative bronchopneumonia, elevated serum IL-1α, IL-6 and G-CSF, and mild dermatitis and/or panniculitis (Extended Data Fig. 6d–f). These data imply that TRIF contributes to inflammation in *Ripk1*^{-/-} mice, but it is ZBP1 that drives perinatal lethality. Accordingly, ZBP1 deficiency alone rescued perinatal lethality in the *Ripk1*^{RHIM/RHIM} mice, whereas TRIF deficiency did not (Fig. 4a–c, Extended Data Fig. 7a). At 7–13 weeks of age, *Ripk1*^{RHIM/RHIM} *Zbp1*^{-/-} mice exhibited mild foci of neutrophilic inflammation that varied in location between animals. In addition, a small amount of RIPK3 autophosphorylation was detected in the dermis (Fig. 4d) and bone-marrow-derived macrophages (BMDMs; Extended Data Fig. 7b). However, RIPK3 autophosphorylation was not detected in *Ripk1*^{RHIM/RHIM} *Zbp1*^{-/-} *Trif*^{-/-} BMDMs, suggesting that the RIPK1 RHIM also suppresses TRIF-dependent RIPK3 activation and inflammation.

We sought biochemical evidence for TRIF or ZBP1 binding to RIPK3 in *Ripk1*^{RHIM/RHIM} macrophages, but found that *Ripk1*^{RHIM/RHIM} fetal liver cells cultured with macrophage colony-stimulating factor yielded approximately fivefold fewer Mac-1⁺ F4/80⁺ macrophages than their wild-type counterparts (Extended Data Fig. 7c). Furthermore, *Ripk1*^{-/-} or *Ripk1*^{RHIM/RHIM} macrophages contained abnormally low amounts of ZBP1, RIPK3 and MLKL (Extended Data Fig. 7d), suggesting selection for cells that were unable to die owing to loss of these proteins. To circumvent these issues, we analysed *Ripk1*^{RHIM/RHIM} *Mkl1*^{-/-} BMDMs. No interaction between RIPK3 and TRIF was detected, but RIPK3 did coimmunoprecipitate with ZBP1 (Fig. 4e). No RIPK3/ZBP1 interaction was detected in wild-type, *Mkl1*^{-/-} or *Ripk1*^{RHIM/RHIM} *Ripk3*^{RHIM/RHIM} BMDMs, indicating that ZBP1 engages the RIPK3 RHIM only when the RIPK1 RHIM is mutated. We explored whether RIPK1 might sequester ZBP1 in wild-type cells and thereby prevent it from interacting with

RIPK3, but failed to detect a RIPK1/ZBP1 interaction. Therefore, how the RIPK1 RHIM suppresses the ZBP1/RIPK3 interaction remains enigmatic.

ZBP1 has two N-terminal Z-DNA binding domains in addition to its C-terminal RHIMs²⁸, and promotes necroptosis in response to murine cytomegalovirus⁷ or influenza⁸ infection. However, it is unclear if ZBP1 acts as a DNA sensor during infection. ZBP1 is also a type-I-interferon-inducible gene²⁹. Interestingly, ZBP1 was more abundant in E18.5 *Ripk1*^{RHIM/RHIM} skin and there was a slight increase in *Ifnb1* expression (Extended Data Fig. 8a, b). However, ZBP1 and *Ifnb1* were not increased in *Ripk1*^{RHIM/RHIM} *Ripk3*^{-/-} or *Ripk1*^{RHIM/RHIM} *Mkl1*^{-/-} skin. It is possible that *Ripk1*^{RHIM/RHIM} skin contained more ZBP1 because of infiltrating leukocytes. Immune cells undergoing necroptosis may also have enhanced RIPK3 autophosphorylation in *Ripk1*^{RHIM/RHIM} skin compared to *Ripk1*^{RHIM/RHIM} *Mkl1*^{-/-} skin (Fig. 1c, Extended Data Fig. 8a).

Collectively, our analyses have revealed an unexpected role for ZBP1 in triggering necroptosis in the perinatal period. Future studies will need to address if ZBP1 senses viral infections and/or DNA in this context, or if mutation of the RIPK1 RHIM is necessary and sufficient for ZBP1/RIPK3 interactions that induce necroptosis (Extended Data Fig. 9). The mechanism by which the RIPK1 RHIM prevents ZBP1 from binding to RIPK3 also warrants further study.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 June; accepted 25 October 2016.

Published online 7 November 2016.

- Rickard, J. A. *et al.* RIPK1 regulates RIPK3–MLKL-driven systemic inflammation and emergency hematopoiesis. *Cell* **157**, 1175–1188 (2014).
- Kaiser, W. J. *et al.* RIP1 suppresses innate immune necrotic as well as apoptotic cell death during mammalian parturition. *Proc. Natl Acad. Sci. USA* **111**, 7753–7758 (2014).
- Dillon, C. P. *et al.* RIPK1 blocks early postnatal lethality mediated by caspase-8 and RIPK3. *Cell* **157**, 1189–1202 (2014).
- Newton, K. *et al.* Activity of protein kinase RIPK3 determines whether cells die by necroptosis or apoptosis. *Science* **343**, 1357–1360 (2014).
- Polykratis, A. *et al.* Cutting edge: RIPK1 kinase inactive mice are viable and protected from TNF-induced necroptosis *in vivo*. *J. Immunol.* **193**, 1539–1543 (2014).
- Chen, W. *et al.* Diverse sequence determinants control human and mouse receptor interacting protein 3 (RIP3) and mixed lineage kinase domain-like (MLKL) interaction in necroptotic signaling. *J. Biol. Chem.* **288**, 16247–16261 (2013).
- Upton, J. W., Kaiser, W. J. & Mocarski, E. S. DAI/ZBP1/DLM-1 complexes with RIP3 to mediate virus-induced programmed necrosis that is targeted by murine cytomegalovirus vIRA. *Cell Host Microbe* **11**, 290–297 (2012).
- Kuriakose, T. *et al.* ZBP1/DAI is an innate sensor of influenza virus triggering the NLRP3 inflammasome and programmed cell death pathways. *Sci. Immunol.* **1**, aag2045 (2016).
- Dannappel, M. *et al.* RIPK1 maintains epithelial homeostasis by inhibiting apoptosis and necroptosis. *Nature* **513**, 90–94 (2014).
- Takahashi, N. *et al.* RIPK1 ensures intestinal homeostasis by protecting the epithelium against apoptosis. *Nature* **513**, 95–99 (2014).
- Kearney, C. J., Cullen, S. P., Clancy, D. & Martin, S. J. RIPK1 can function as an inhibitor rather than an initiator of RIPK3-dependent necroptosis. *FEBS J.* **281**, 4921–4934 (2014).
- Orozco, S. *et al.* RIPK1 both positively and negatively regulates RIPK3 oligomerization and necroptosis. *Cell Death Differ.* **21**, 1511–1521 (2014).
- Newton, K. & Manning, G. Necroptosis and inflammation. *Ann. Rev. Biochem.* **85**, 743–763 (2016).
- Sun, X., Yin, J., Starovasnik, M. A., Fairbrother, W. J. & Dixit, V. M. Identification of a novel homotypic interaction motif required for the phosphorylation of receptor-interacting protein (RIP) by RIP3. *J. Biol. Chem.* **277**, 9505–9511 (2002).
- Cho, Y. S. *et al.* Phosphorylation-driven assembly of the RIP1–RIP3 complex regulates programmed necrosis and virus-induced inflammation. *Cell* **137**, 1112–1123 (2009).
- He, S. *et al.* Receptor interacting protein kinase-3 determines cellular necrotic response to TNF-α. *Cell* **137**, 1100–1111 (2009).
- Zhang, D. W. *et al.* RIP3, an energy metabolism regulator that switches TNF-induced cell death from apoptosis to necrosis. *Science* **325**, 332–336 (2009).

18. Kelliher, M. A. *et al.* The death domain kinase RIP mediates the TNF-induced NF- κ B signal. *Immunity* **8**, 297–303 (1998).
19. Newton, K. *et al.* Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. *Cell* **134**, 668–678 (2008).
20. Ea, C. K., Deng, L., Xia, Z. P., Pineda, G. & Chen, Z. J. Activation of IKK by TNF α requires site-specific ubiquitination of RIP1 and polyubiquitin binding by NEMO. *Mol. Cell* **22**, 245–257 (2006).
21. Bertrand, M. J. *et al.* cIAP1 and cIAP2 facilitate cancer cell survival by functioning as E3 ligases that promote RIP1 ubiquitination. *Mol. Cell* **30**, 689–700 (2008).
22. Varfolomeev, E. *et al.* c-IAP1 and c-IAP2 are critical mediators of tumor necrosis factor alpha (TNF α)-induced NF- κ B activation. *J. Biol. Chem.* **283**, 24295–24299 (2008).
23. Ermolaeva, M. A. *et al.* Function of TRADD in tumor necrosis factor receptor 1 signaling and in TRIF-dependent inflammatory responses. *Nat. Immunol.* **9**, 1037–1046 (2008).
24. Pobeinskaya, Y. L. *et al.* The function of TRADD in signaling through tumor necrosis factor receptor 1 and TRIF-dependent Toll-like receptors. *Nat. Immunol.* **9**, 1047–1054 (2008).
25. Varfolomeev, E. *et al.* IAP antagonists induce autoubiquitination of c-IAPs, NF- κ B activation, and TNF α -dependent apoptosis. *Cell* **131**, 669–681 (2007).
26. Wu, X. N. *et al.* Distinct roles of RIP1–RIP3 hetero- and RIP3–RIP3 homo-interaction in mediating necroptosis. *Cell Death Differ.* **21**, 1709–1720 (2014).
27. Salmena, L. & Hakem, R. Caspase-8 deficiency in T cells leads to a lethal lymphoinfiltrative immune disorder. *J. Exp. Med.* **202**, 727–732 (2005).
28. Kaiser, W. J., Upton, J. W. & Mocarski, E. S. Receptor-interacting protein homotypic interaction motif-dependent control of NF- κ B activation via the DNA-dependent activator of IFN regulatory factors. *J. Immunol.* **181**, 6427–6434 (2008).
29. Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501–505 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Diaz, M. Long, C. Allen, M. Garcia, A. Verducci and J. Anunciacion for animal husbandry, R. Newman, K. Heger and the Genentech genetic analysis and histology laboratories for technical assistance, and J. Lai and S. Stawicki for antibody project management.

Author Contributions M.R.G. and S.W. generated the *Ripk1*^{RHIM/+} and *Ripk3*^{RHIM/+} mice, K.N., A.M., D.L.D., K.E.W., and A.S. designed and performed experiments, V.C.P. and J.R.L. characterized the RIPK1 and RIPK3 autophosphorylation sites, M.S. performed immunohistochemistry with quantitative analyses by H.N., J.D.W. analysed histological data, and V.M.D. helped with experimental design.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.M.D. (dixit@gene.com) or K.N. (knewton@gene.com).

Reviewer Information *Nature* thanks H. Walczak and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Mice. The experiments were not randomized and the investigators were not blinded to allocations during experiments and outcome assessment. No statistical methods were used to predetermine sample size.

Ripk3^{−/−} (ref. 4), *Ripk3*^{D161N/D161N} (ref. 4), *Ripk1*^{D138N/D138N} (ref. 4), *Ripk1*^{−/−} (ref. 4), *Casp8*^{−/−} (ref. 4), *Tradd*^{−/−} (ref. 22), *Mkl1*^{−/−} (ref. 30), *Zbp1*^{−/−} (ref. 31), and *Trif*^{−/−} (ref. 32) mice were described previously. A second *Zbp1*^{−/−} strain generated by Taconic (Germany) using C57BL/6N^{Tac} ES cells was also crossed to *Ripk1*^{RHIM/RHIM} mice and rescued perinatal lethality similar to the published *Zbp1*^{−/−} strain³¹, so results were pooled. The Taconic strain lacked the *Zbp1* 5' UTR and exon 1 corresponding to NCBI37/mm9 chr2:173,043,537–173,045,687. Loss of ZBP1 protein was confirmed in macrophages (Extended Data Fig. 7b). *Ripk1*^{RHIM/+} and *Ripk3*^{RHIM/+} mice were generated at Genentech using C57BL/6N C2 ES cells. A FRT-flanked neomycin selection cassette inserted upstream of *Ripk1* exon 10 at position chr13:34,029,090 (GRCm38/mm10 assembly) or upstream of *Ripk3* exon 10 at position chr14:55,785,501 (reverse strand) was deleted from the targeted embryonic stem cells before microinjection by adenoviral delivery of Flpe. *Ripk1* sequence 5'-ATTGAGATTGGA-3' encoding RHIM residues IQIG was replaced with the sequence 5'-GCTGCGGCTGCA-3'. *Ripk3* sequence 5'-GTGAGATTGGG-3' encoding RHIM residues VQIG was replaced with the sequence 5'-GCAGCCGCGGCT-3'.

Ripk1^{RHIM} genotyping primers 5'-CCACATTCTTGCCAACACTG-3' and 5'-GCAAGTATTGTTGGTGGTGG-3' amplified 299 base pair (bp) wild-type and 333-bp knock-in DNA fragments. *Ripk3*^{RHIM} genotyping primers 5'-AGCAGGCACTACTCTTTGAGCT-3' and 5'-CTGTGCTTGCTCATCTTGGC-3' amplified 325-bp wild-type and 359-bp knock-in DNA fragments. *Zbp1* genotyping primers 5'-AGACATTAGAAAGCACAGATC-3', 5'-TGGCCTCTCCTTCATTCC-3' and 5'-CTCCTAGGTCACTGACTCTC amplified 145-bp wild-type and 294-bp knockout (Taconic strain) DNA fragments.

For timed pregnancies, mice were designated E0.5 on the morning a vaginal plug was detected. When determining offspring numbers, pups were clipped between 4–6 days of age. The Genentech institutional animal care and use committee approved all protocols.

Immunohistochemistry and TUNEL staining. Formalin-fixed, paraffin-embedded tissue sections were stained with 5 µg ml^{−1} rabbit anti-mouse phospho-RIPK3 antibody (GEN135-35-9, Genentech) recognizing phosphorylated residues Thr231, Ser232 or 0.05 µg ml^{−1} rabbit anti-cleaved caspase-3 (CC3) antibody (Cell Signaling Technology). Immunohistochemistry was performed on the Ventana Discovery XT platform with CC1 standard antigen retrieval. The reaction was detected with the HQ amplification system (p-RIPK3) or OmniMap detection system (CC3) using DAB as the chromogen and haematoxylin counterstain. TUNEL staining was performed using the Apoptag kit (EMD Millipore) with proteinase K digestion.

Stained slides were digitized using a Nanozoomer digital slide scanner. Regions of interest in the dorsal skin and thymus were manually annotated on the digital images and Matlab software (MathWorks) was used to quantify the total region of interest tissue area and the immunolabelled area. Final results were expressed as the percent immunolabelled area per total region of interest tissue area.

Cell culture. Primary MEFs isolated from E13.5 or E14.5 embryos were grown in the high glucose version of Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% heat inactivated-fetal bovine serum, 2 mM glutamine, 10 mM HEPES (pH 7.2), 1 × non-essential amino acids solution, 100 U ml^{−1} penicillin, and 100 µg ml^{−1} streptomycin on tissue culture dishes pre-coated with 0.1% gelatin in PBS. Primary macrophages were differentiated from bone marrow cells in non-treated plates using the same medium supplemented with 25 ng ml^{−1} M-CSF (R&D Systems) for 5–6 days. Bone-marrow-derived macrophages (BMDMs) were then collected and replated for experiments. BMDMs and MEFs were stimulated with 100 ng ml^{−1} murine TNF (Genentech), 20 µM Z-VAD-FMK (Promega), 100 ng ml^{−1} ultra-pure LPS-EB (Invivogen), 50 µg ml^{−1} LMW poly I-C (Invivogen), 2 µM BV6 (Genentech), 1 µg ml^{−1} TNF-Flag (Enzo Life Sciences) or 20 µg ml^{−1} cycloheximide (Sigma). BMDM viability was assessed after YOYO-1 (Molecular Probes) staining and live-cell imaging in an Incucyte system (Essen Bioscience). Alternatively, BMDMs were collected from non-treated plates with a cell scraper into the culture medium along with any floating dead cells, stained with 2.5 µg ml^{−1} propidium iodide (PI; BD Biosciences), and analysed in a FACSCanto II (BD Biosciences).

E14.5 fetal liver cells were plated overnight in high glucose DMEM supplemented with 10% heat inactivated-fetal bovine serum, 2 mM glutamine, 100 µM

asparagine, 55 µM 2-mercaptoethanol, 50 U ml^{−1} penicillin, and 50 µg ml^{−1} streptomycin. 2.5 × 10⁶ viable, nucleated, non-adherent cells were then plated on a 15-cm non-treated dish in 50 ng ml^{−1} M-CSF (R&D Systems) for 7 days. Adherent cells were collected with a cell scraper, counted, and then analysed by flow cytometry after staining with antibodies recognizing FITC-conjugated MCA497, CI:A3-1 anti-F4/80 (Bio-Rad), PE-conjugated M1/70 anti-Mac-1 (BD Biosciences), and APC-conjugated RB6-8C5 anti-Gr-1 (BD Biosciences). Dead cells that stained with PI were excluded from analyses.

Lymph node cells were stained with FITC-conjugated 145-2C11 anti-CD3 (BD Biosciences) and APC-conjugated RA3-6B2 anti-B220 (BD Biosciences) antibodies for flow cytometric analysis. Dead cells that stained with PI were excluded from analyses.

To test the specificity of the GEN135-35-9 anti-phospho-RIPK3 antibody, 293T cells (ATCC CRL-3216) were transfected with N-terminal 3 × Flag-tagged mouse RIPK3 variants in vector pCMV-3Tag-6 (Agilent) using lipofectamine 2000 (ThermoFisher). The cell line was neither authenticated nor tested for mycoplasma.

Immunoprecipitation and western blotting. Cells were lysed in 20 mM Tris, HCl pH 7.5, 135 mM NaCl, 1.5 mM MgCl₂, 1 mM EGTA, 1% Triton X-100, 10% glycerol, phosphatase inhibitor (Roche), and complete protease inhibitor cocktail (Roche) (Figs 2, 4e, Extended Data Figs 1c, 3g–h, 4a–c, 5 and 8a). Insoluble material was removed by centrifugation at 20,000g before immunoprecipitation or addition of LDS sample buffer. Alternatively, cells were lysed in 10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2.5 mM MgCl₂, 0.5 mM CaCl₂, 1% NP40, phosphatase/protease inhibitors (Roche) and DNase (~80 U ml^{−1}, Qiagen) (Extended Data Figs 1d, 4d, 7). The whole-cell lysate was denatured directly in LDS sample buffer.

Immunoprecipitating antibodies recognized RIPK1 (610459, BD Biosciences), RIPK3 (NBP1-77299, Novus Biologicals), ZBP1 (clone Zippy-1, Adipogen), Flag (clone M2, Sigma), or were irrelevant control IgGs (Mouse IgG2a, BD Biosciences, 553454; or Rabbit IgG, Millipore, 12-370). Antibody complexes were recovered with magnetic protein A/G beads (Pierce) or magnetic Flag beads (Sigma).

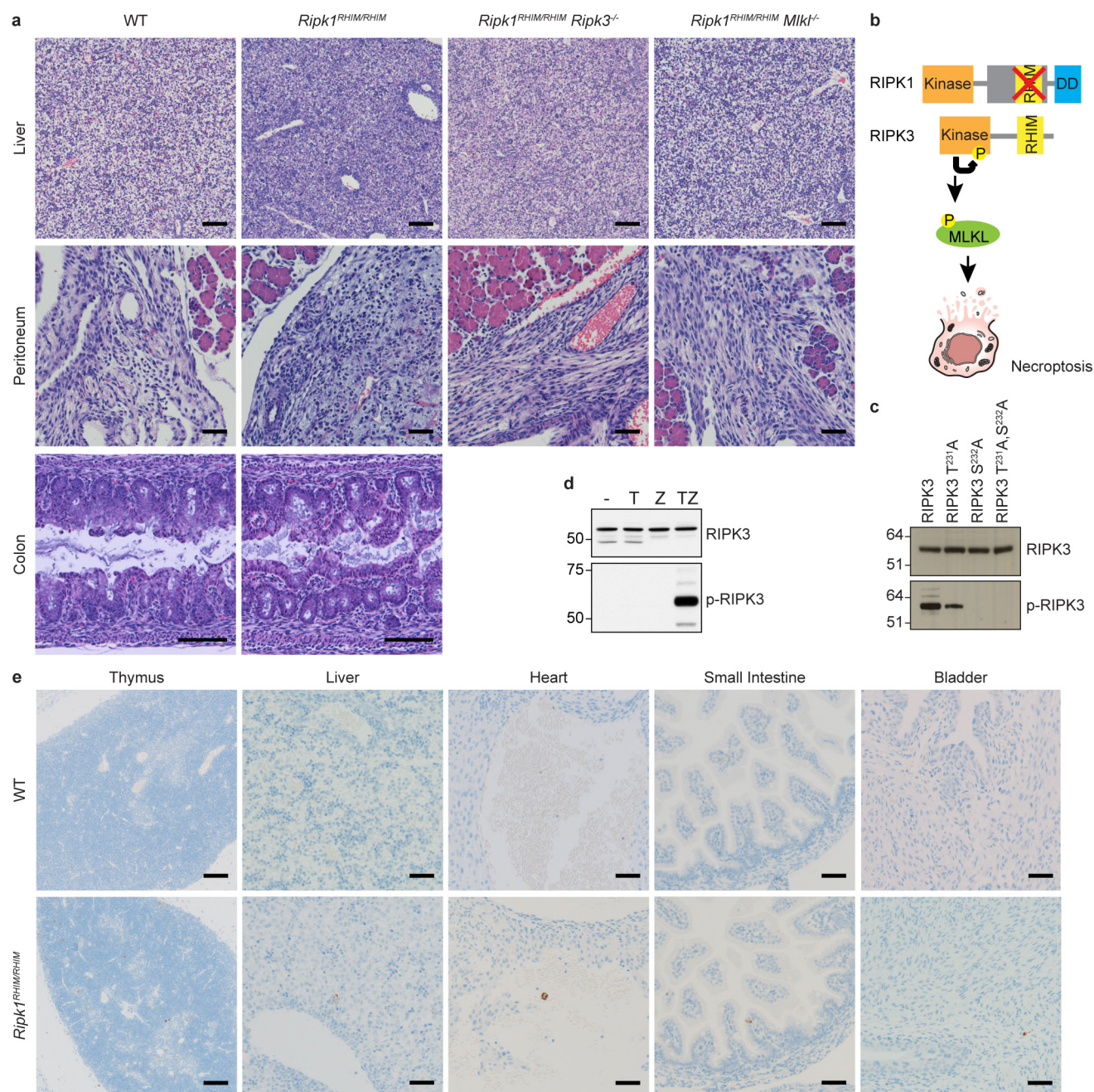
Western blot antibodies recognized RIPK3 (1G6.1.4, Genentech), phosphorylated RIPK3 Thr231, Ser232 (GEN135-35-9 raised against peptide ELVDK(pT) (pS)LLRET, Genentech), RIPK1 (BD Biosciences), phosphorylated RIPK1 Ser166 (GEN150-33-4 raised against peptide GVASFKTW(pS)KLTKKEK, Genentech), Flag (clone M2, Sigma), β-actin (MP Biomedicals, mouse clone C4), FADD (1.28E12, Genentech), caspase-8 (1G12, Enzo Life Sciences), FLIP (2.21H2, Genentech), MLKL (1G12, Genentech), TRAF2 (7187, Santa Cruz Biotechnology), cIAP1 (ALX-803-335-C100, Enzo Life Sciences), TRIF (GENW2-6, Genentech) and TRADD (GEN21-3, Genentech). The following antibodies were from Cell Signaling Technologies: p-ERK (9101), ERK (9102), p-JNK (4668), JNK (9258), p-IκBα (2859), IκBα (9242), p-p65/RelA (3033), p65/RelA (8242), p-p38 (9211), and p38 (8690). RIPK3 antibody 1G6 was biotinylated using a Biotin-xx Microscale Protein Labelling Kit (Molecular Probes) in order to detect RIPK3 in ZBP1 immunoprecipitates.

Chemokines and cytokines. E18.5 mouse skin was homogenized in ice-cold 50 mM Tris HCl (pH 7.4), 150 mM NaCl, 2 mM EDTA, 0.5% sodium deoxycholate, 0.1% SDS, 1% NP-40, and 1 mM DTT supplemented with complete protease inhibitor cocktail (Roche). Insoluble material was pelleted at 20,000xg and the protein concentration in the soluble fraction measured by Bio-rad Protein Assay. Skin lysates adjusted to 1 mg protein/ml or sera were analysed by Bio-Plex Pro Mouse Cytokine 23-plex (Bio-Rad).

Quantitative reverse-transcription PCR. RNA was isolated from E18.5 skin with an RNeasy mini-kit (Qiagen). An on-column DNase treatment was included. cDNA was generated from each RNA sample using a High Capacity cDNA Reverse Transcription kit (Applied Biosystems). Gene expression assays were from Life Technologies: *Ifnb1*: Mm00439552_s1; *Gapdh*: 4352339E. *Ifnb1* levels were normalized against *Gapdh*.

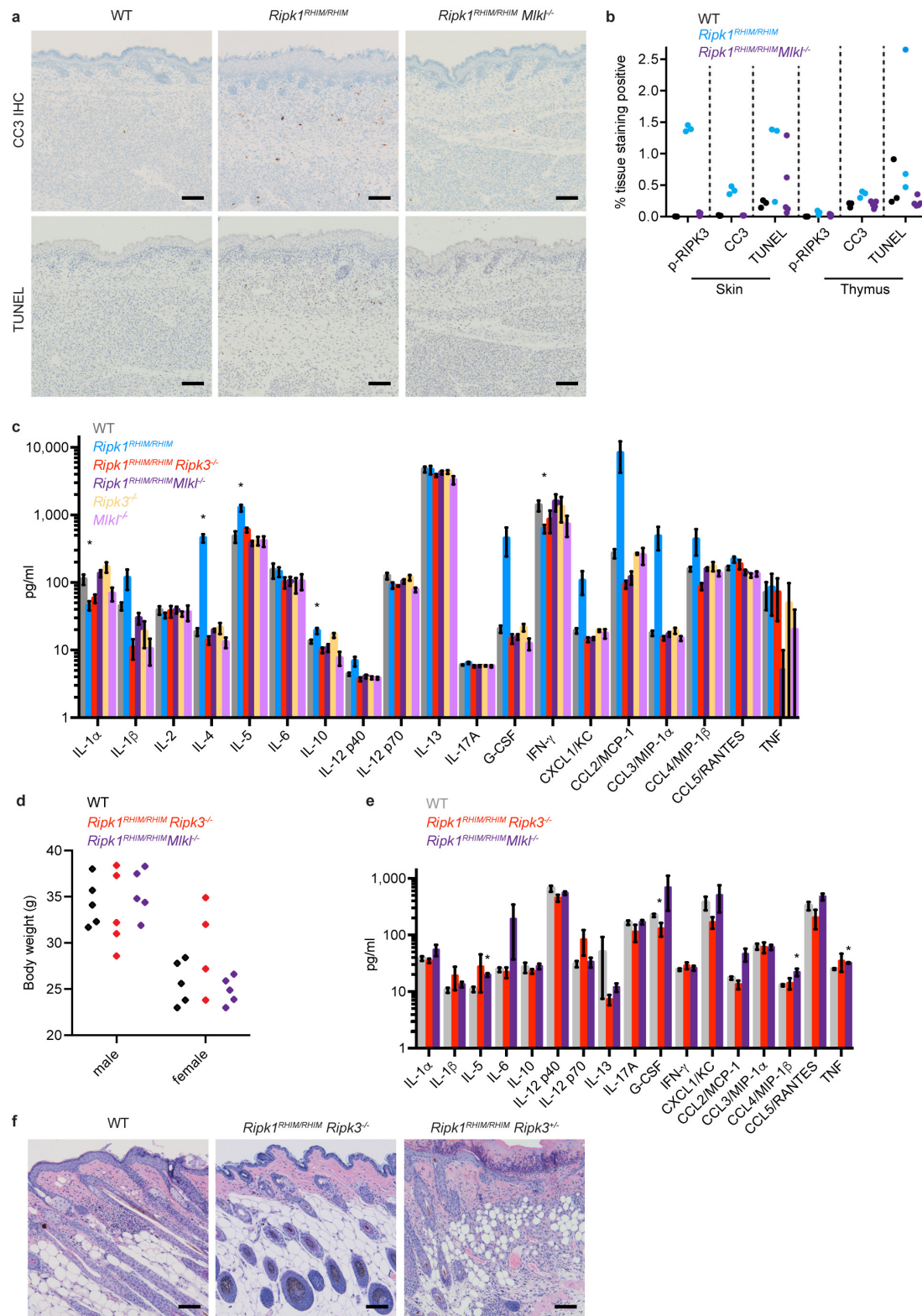
Data availability. The data sets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

- Murphy, J. M. *et al.* The pseudokinase MLKL mediates necroptosis via a molecular switch mechanism. *Immunity* **39**, 443–453 (2013).
- Ishii, K. J. *et al.* TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines. *Nature* **451**, 725–729 (2008).
- Yamamoto, M. *et al.* Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science* **301**, 640–643 (2003).



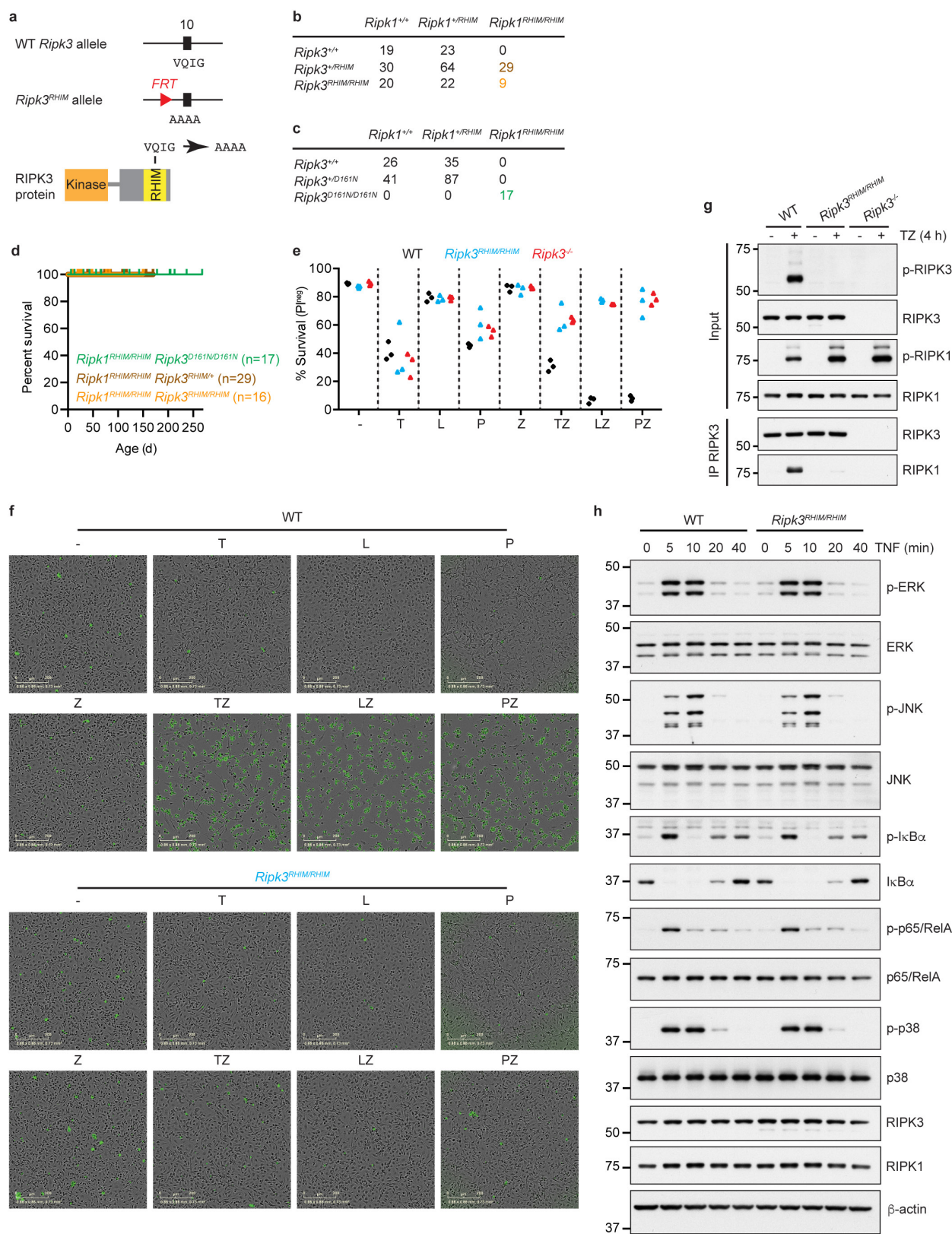
Extended Data Figure 1 | Characterization of *Ripk1* mutant mice with a monoclonal antibody recognizing autophosphorylated RIPK3. **a**, E18.5 peritoneum and liver sections or E17.5 colon sections stained with H&E. Scale bar, 50 μ m (peritoneum) or 100 μ m (colon and liver). Results are representative of 3 (wild type, *Ripk1*^{RHIM/RHIM} and *Ripk1*^{RHIM/RHIM} *Ripk3*^{-/-}) or 5 embryos (*Ripk1*^{RHIM/RHIM} *Mlki*^{-/-}). **b**, Model for RIPK3 activation following mutation of the RIPK1 RHIM. **c**, Western blots of 293T cells transfected with Flag-tagged mouse RIPK3 variants,

overexpression being sufficient to activate RIPK3 autophosphorylation on Thr231 and Ser232 based on mass spectrometry (data not shown). p-RIPK3, RIPK3 phosphorylated on Thr231, Ser232. **d**, Western blots of macrophages at 4 h after treatment. T, TNF. Z, Z-VAD-FMK. Results are representative of 5 independent experiments. **e**, E18.5 embryo sections. Cells containing p-RIPK3 are stained brown. Scale bar, 50 μ m (except thymus, which has a 100 μ m scale bar). Results are representative of 3 mice of each genotype. For gel source data, see Supplementary Fig. 1.



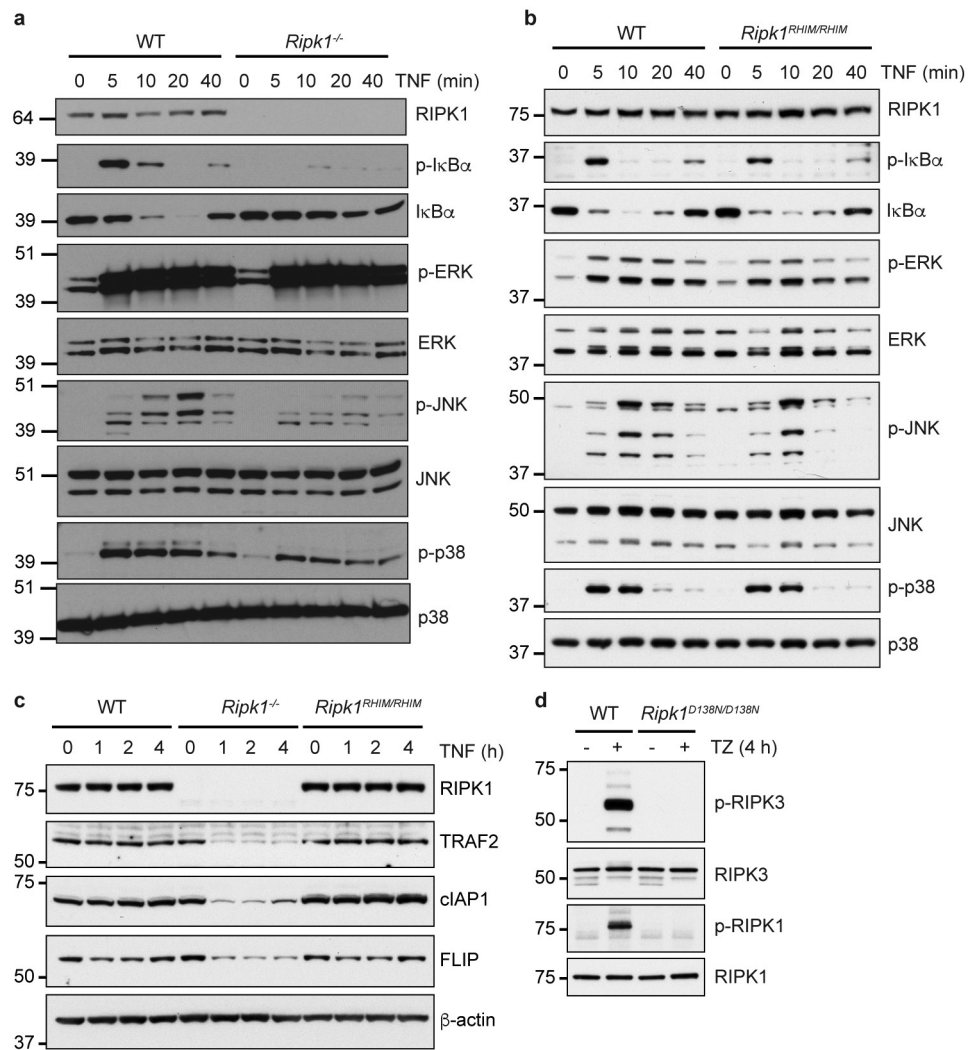
Extended Data Figure 2 | Characterization of *Ripk1^{RHIM/RHIM}* mice lacking RIPK3 or MLKL. **a**, E18.5 skin sections stained for cleaved caspase-3 (CC3) by immunohistochemistry (IHC) or nicked DNA by TUNEL assay. WT, wild-type. Scale bar, 100 μ m. **b**, Graph indicates the percentage of tissue that was positive for RIPK3 phosphorylated on Thr231, Ser232 (p-RIPK3), CC3, or TUNEL in E18.5 thymus and skin. Tissue sections from 3 (wild type and *Ripk1^{RHIM/RHIM}*) or 5 (*Ripk1^{RHIM/RHIM} Mlkl^{-/-}*) embryos were analysed. Each symbol represents one embryo. **c**, E18.5 skin cytokines and chemokines. Bars indicate the mean \pm s.e.m. of 5 embryos of each genotype. * $P < 0.05$ when comparing wild-type and *Ripk1^{RHIM/RHIM}* embryos by 2-tailed t -test.

d, Body weights of mice aged 7–8 months. Each symbol represents one mouse. **e**, Serum cytokines and chemokines of mice aged 8–12 months. Bars indicate the mean \pm s.e.m. of 3 females and 3 males of each genotype. * $P < 0.05$ by 2-tailed t -test when compared to wild-type. **f**, H&E stained skin sections of female mice aged 9 months. Scale bar, 100 μ m. Note that mild dermatitis and/or panniculitis were observed in 3 out of 5 wild-type mice aged 8–10 months. More severe dermatitis was seen in 1 out of 3 *Ripk1^{RHIM/RHIM} Ripk3^{-/-}* mice aged 9 months, and 2 out of 4 *Ripk1^{RHIM/RHIM} Mlkl^{-/-}* mice aged 9 months. Given that dermatitis is a common finding in older C57BL/6 mice, larger cohorts would be needed to exclude age-related variability.

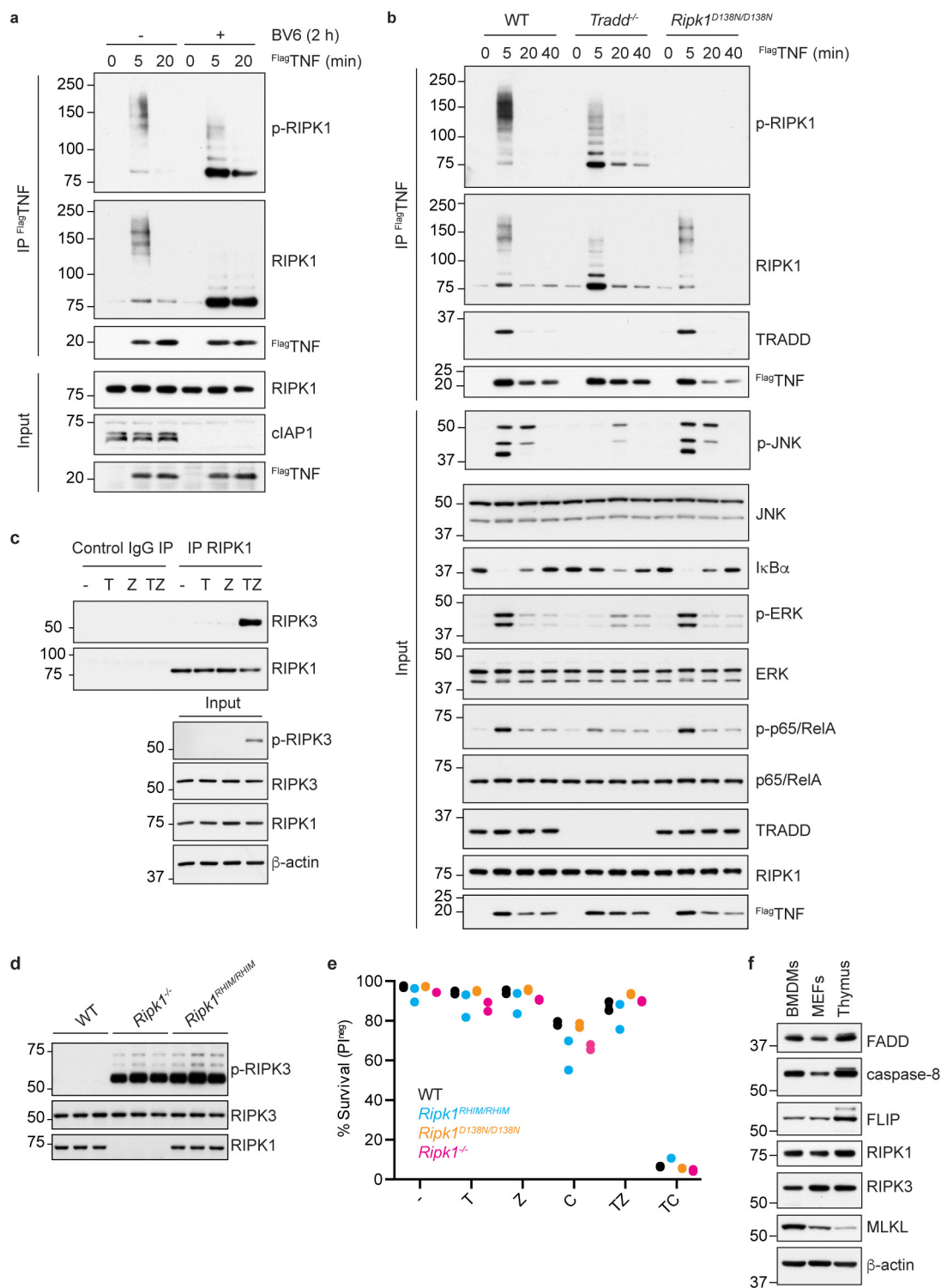


Extended Data Figure 3 | *Ripk1*^{*RHIM/RHIM*} mice expressing catalytically inactive or RHIM mutant RIPK3 are viable. **a**, Organisation of the *Ripk3*^{*RHIM*} mutant allele. Black boxes indicate exons. RHIM, RIP homotypic interaction motif. **b**, **c**, Numbers of offspring from intercrosses of compound heterozygote parents. **d**, Kaplan-Meier plot of mouse survival. Note that seven *Ripk1*^{*RHIM/RHIM*}*Ripk3*^{*RHIM/RHIM*} mice were obtained from intercrossing *Ripk1*^{*RHIM/+*}*Ripk3*^{*RHIM/RHIM*} and *Ripk1*^{*RHIM/RHIM*}*Ripk3*^{*RHIM/RHIM*} mice and are not included in **b**, **e**, Graph indicates the percentage of macrophages that are viable and not stained

by propidium iodide (PI^{ncf}) at 16 h after treatment. T, TNF; L, LPS; P, poly I-C; Z, Z-VAD-FMK. Symbols represent cells from different mice. **f**, Representative images of macrophages stained with YOYO-1 at 16 h after the same treatments as in **e**. Note that this assay does not reveal death in response to TNF, LPS or poly I-C individually. Therefore, the death quantified in **e** due to TNF, LPS or poly I-C individually is probably linked to mechanical scraping of the cells before flow cytometry. **g**, **h**, Western blots of macrophages. Results are representative of 2 independent experiments. For gel source data, see Supplementary Fig. 1.

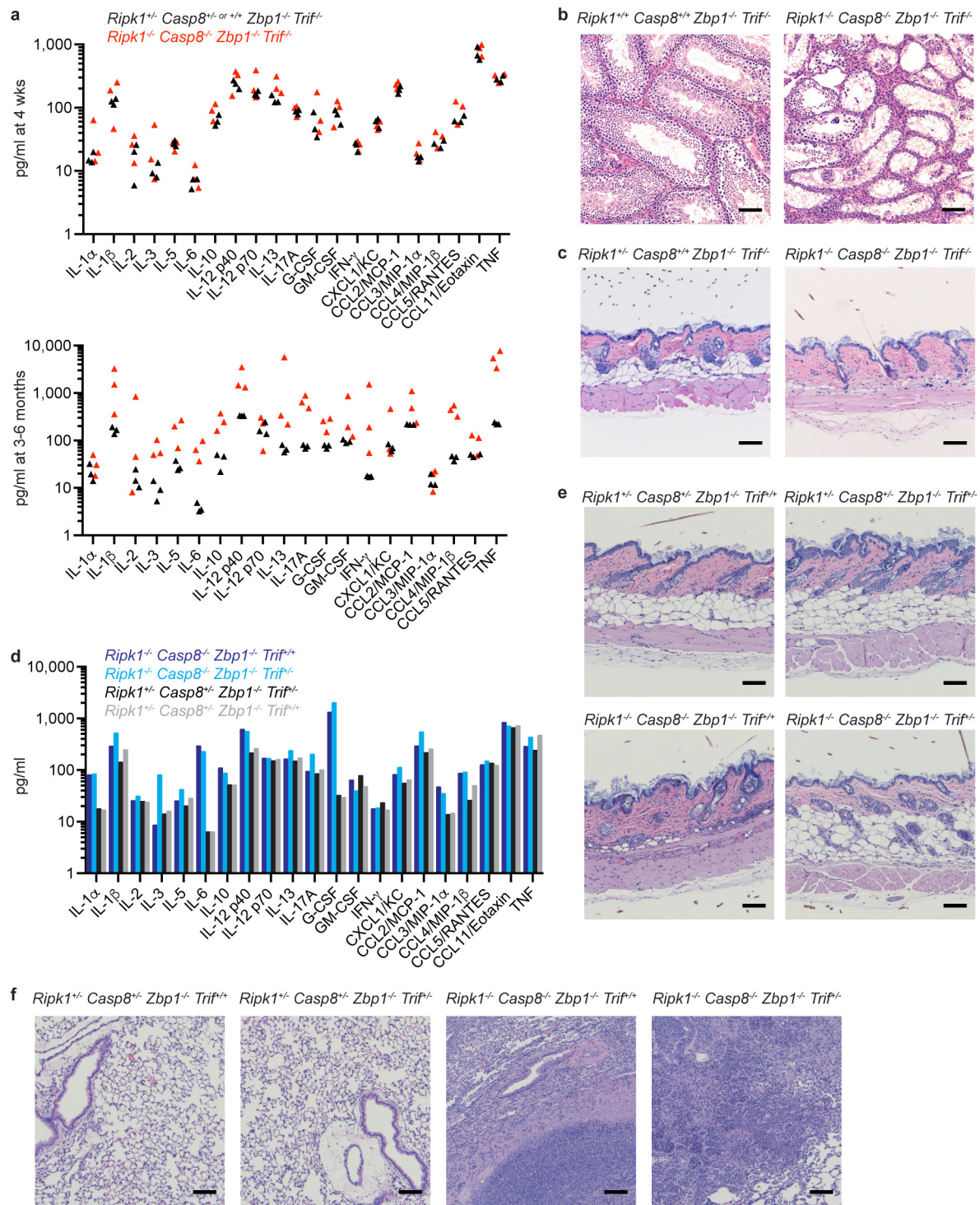


Extended Data Figure 4 | Comparison of *Ripk1*^{-/-} and *Ripk1*^{RHIM/RHIM} MEFs. a–d, Western blots of MEFs (a–c) or macrophages (d). ‘p-’ indicates an active, phosphorylated variant of the protein. Results are representative of 2 (b), 3 (a) or 5 (d) independent experiments. For gel source data, see Supplementary Fig. 1.



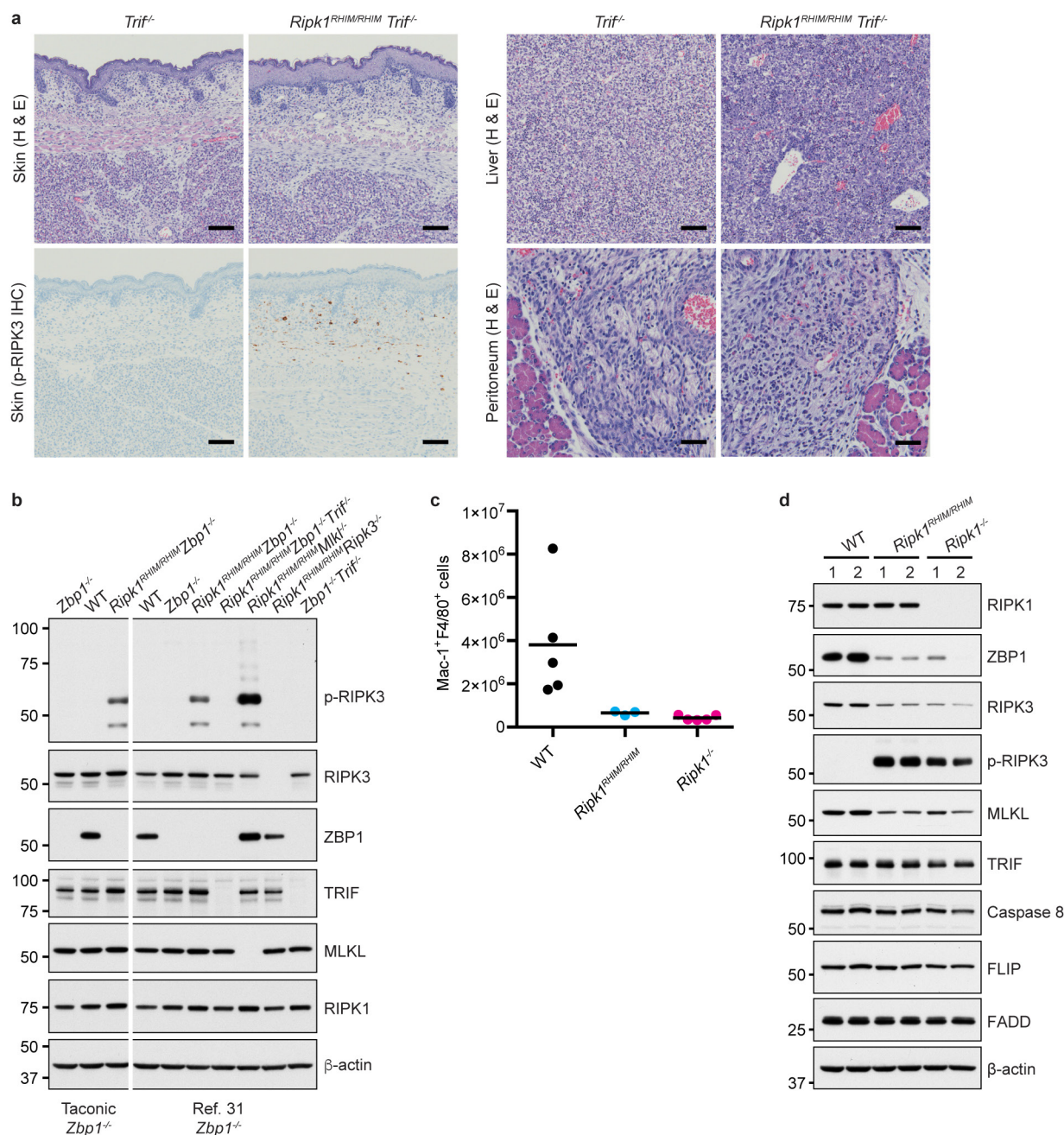
Extended Data Figure 5 | Biochemical analyses of RIPK1 following TNF stimulation. **a–d**, Western blots of MEFs (**a**, **c**, **d**) or macrophages (**b**). MEFs in **a** that received BV6 were pretreated for 2 h before stimulation with TNF-Flag. In **d**, MEFs derived from 3 different embryos of each genotype were analysed. **e**, Graph indicates the percentage of primary MEFs that are viable and not stained by propidium iodide (PI^{nes}) at 25 h

after treatment. T, TNF; Z, Z-VAD-FMK. C, cycloheximide. Each symbol represents cells from a different embryo. **f**, Western blots of wild-type BMDMs, MEFs, and adult mouse thymus. Results are representative of 2 independent experiments. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 6 | Characterization of *Ripk1^{-/-} Casp8^{-/-} Trif^{-/-} Zbp1^{-/-}* mice. **a**, Serum cytokines and chemokines of mice aged 4 weeks (upper graph) or 3–6 months (lower graph). Each symbol represents one mouse. **b**, H&E stained testes sections from mice aged 12 weeks. Scale bar, 100 μ m. Results are representative of two

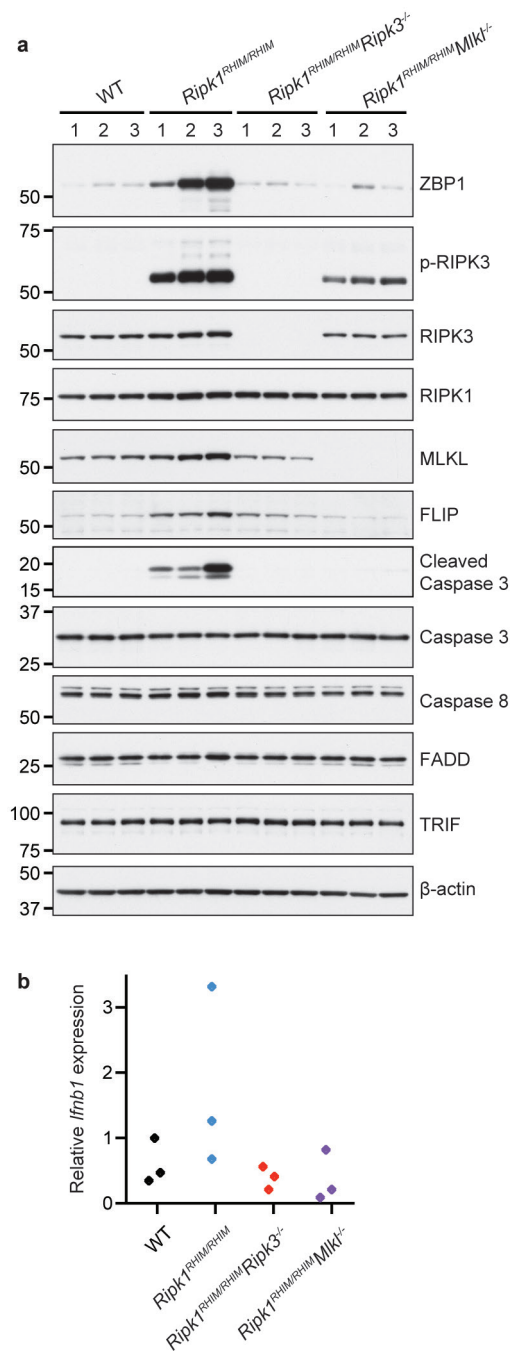
Ripk1^{-/-} Casp8^{-/-} Trif^{-/-} Zbp1^{-/-} males. **c**, H&E stained skin sections from female littermates aged 19 weeks. Scale bar, 100 μ m. Results are representative of three *Ripk1^{-/-} Casp8^{-/-} Trif^{-/-} Zbp1^{-/-}* mice. **d**, Serum cytokines and chemokines of mice aged 7 weeks. **e**, **f**, H&E stained skin (**e**) and lung (**f**) sections from the mice in **d**. Scale bar, 100 μ m.



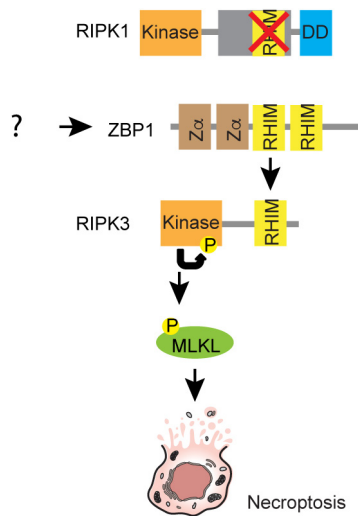
Extended Data Figure 7 | The effect of ZBP1 and/or TRIF deficiency on *Ripk1*^{RHIM/RHIM} mice. **a**, E18.5 embryo sections stained with H&E or an antibody recognizing RIPK3 autophosphorylation on Thr231 and Ser232 (p-RIPK3 IHC). Scale bar, 50 μ m (peritoneum) or 100 μ m (skin and liver). Results are representative of 2 *Trif*^{-/-} and 3 *Ripk1*^{RHIM/RHIM}*Trif*^{-/-} embryos. **b**, Western blots of bone marrow-derived macrophages. Results are representative of 2 independent experiments. **c**, Graph indicates the

number of F4/80⁺Mac-1⁺ macrophages obtained after 2.5×10^6 E14.5 viable, nucleated fetal liver cells were cultured in M-CSF for 7 days. Each symbol represents cells from one embryo. Lines indicate the mean.

d, Western blot analysis of the macrophages in **c**. Cells from two embryos of each genotype were analysed. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 8 | Characterization of E18.5 *Ripk1^{RHIM/RHIM}* skin. **a**, Western blot analysis of E18.5 skin. Three embryos of each genotype were analysed. For gel source data, see Supplementary Fig. 1. **b**, Graph indicates relative *Ifnb1* gene expression in E18.5 skin. Each symbol represents one embryo.



Extended Data Figure 9 | Model for activation of RIPK3 by ZBP1 in the perinatal period. Mutation of the RIPK1 RHIM promotes RHIM-dependent interactions between ZBP1 and RIPK3, resulting in RIPK3 autophosphorylation and MLKL-dependent necroptosis. It is unclear if the ZBP1/RIPK3 interaction is driven by an upstream stimulus, such as the binding of RNA or DNA to the Z α domains in ZBP1.

The SND proteins constitute an alternative targeting route to the endoplasmic reticulum

Naama Aviram¹, Tslil Ast^{1†}, Elizabeth A. Costa², Eric C. Arakel³, Silvia G. Chuartzman¹, Calvin H. Jan², Sarah Haßdenteufel⁴, Johanna Dudek⁴, Martin Jung⁴, Stefan Schorr⁴, Richard Zimmermann⁴, Blanche Schwappach^{3,5}, Jonathan S. Weissman² & Maya Schuldiner¹

In eukaryotes, up to one-third of cellular proteins are targeted to the endoplasmic reticulum, where they undergo folding, processing, sorting and trafficking to subsequent endomembrane compartments¹. Targeting to the endoplasmic reticulum has been shown to occur co-translationally by the signal recognition particle (SRP) pathway² or post-translationally by the mammalian transmembrane recognition complex of 40 kDa (TRC40)^{3,4} and homologous yeast guided entry of tail-anchored proteins (GET)^{5,6} pathways. Despite the range of proteins that can be catered for by these two pathways, many proteins are still known to be independent of both SRP and GET, so there seems to be a critical need for an additional dedicated pathway for endoplasmic reticulum relay^{7,8}. We set out to uncover additional targeting proteins using unbiased high-content screening approaches. To this end, we performed a systematic visual screen using the yeast *Saccharomyces cerevisiae*^{9,10}, and uncovered three uncharacterized proteins whose loss affected targeting. We suggest that these proteins work together and demonstrate that they function in parallel with SRP and GET to target a broad range of substrates to the endoplasmic reticulum. The three proteins, which we name Snd1, Snd2 and Snd3 (for SRP-independent targeting), can synthetically compensate for the loss of both the SRP and GET pathways, and act as a backup targeting system. This explains why it has previously been difficult to demonstrate complete loss of targeting for some substrates. Our discovery thus puts in place an essential piece of the endoplasmic reticulum targeting puzzle, highlighting how the targeting apparatus of the eukaryotic cell is robust, interlinked and flexible.

To uncover factors that contribute to endoplasmic reticulum (ER) targeting, we devised a high-content screen in the yeast *S. cerevisiae* (Fig. 1a). We followed the fate of a model substrate, Gas1, which is known to be completely SRP-independent and only partially dependent on GET for targeting to the ER^{9,10}. Using automated techniques^{11,12}, we integrated fluorescently tagged Gas1 (RFP–Gas1) into around 6,000 strains harbouring mutations in every yeast gene^{13,14} and imaged them on an automated fluorescence microscopy platform¹⁵. We visually searched for strains in which Gas1 localized differently from wild-type or negative controls (where it localizes to the cell walls and vacuoles; Fig. 1b, Extended Data Fig. 1a). In strains mutated in the canonical translocon subunit Sec61 (ref. 16), the auxiliary complex that facilitates SRP-independent translocation (Sec62, Sec63, Sec66, Sec72)¹⁷ or the GET pathway (Get3)⁶, Gas1 accumulated in the cytosol, where it aggregated as expected (Fig. 1b) (for a full list of genes that displayed an altered phenotype, see Supplementary Table 1).

Three uncharacterized mutants showed a similar mislocalization effect to the $\Delta get3$ strain (Fig. 1b). Because the Gas1 foci that formed in the three mutants co-localized with a soluble misfolded marker¹⁸ (Extended Data Fig. 1b), and as the mutants did not affect the

localization of an SRP-dependent substrate (Extended Data Fig. 1c), we named these new elements SND (SRP-independent targeting) proteins.

Snd1 (encoded by *YDR186C*, here called *SND1*) is predicted to be soluble (Extended Data Fig. 1d), localizes to the cytosol¹⁹ (Fig. 1c; for verification that all tagged SND proteins are functional see Extended Data Fig. 1e) and was proposed to be a peripheral ribosomal protein²⁰. Snd2 (encoded by *ENV10*, also known as *YLR065C*, here called *SND2*) is predicted to have four transmembrane domains (Extended Data Fig. 1d), localizes to the ER membrane (Fig. 1c) and was previously shown to affect carboxypeptidase Y (CPY) maturation²¹. The human orthologue of Snd2, TMEM208 (here referred to as hSnd2), was previously shown to localize to the ER when tagged²². We confirmed localization of the native protein by raising antibodies against hSnd2 and could detect it in canine pancreatic microsomes and enriched in ER fractions of HEK293 cells (Extended Data Fig. 1f, g). The third protein, Snd3 (encoded by *PHO88*, also known as *YBR106W*, here called *SND3*), is predicted to have one transmembrane domain (Extended Data Fig. 1d), and is localized to the ER¹⁹ (Fig. 1c). Loss of *SND3* has been shown to affect secretion of the yeast acid phosphatase via an unknown mechanism²³.

To investigate whether the SND proteins work cooperatively in a pathway or complex, we immunoprecipitated GFP–Snd2 and Snd3–GFP. Mass spectrometry analysis showed that both of them interacted physically with components of the targeting and translocation apparatus of the cell (Supplementary Table 2). Interestingly, GFP–Snd2 pull-downs were enriched for Snd1 (Supplementary Table 2). Snd2 also co-immunoprecipitated with Snd3 and Sec61 (Fig. 1d). Moreover, we found Snd2 and Snd3 in a complex with the translocon when assayed by blue native gel electrophoresis followed by SDS–PAGE (Extended Data Fig. 2a). In support of the idea that these proteins have a role in targeting substrates to the translocation machinery, we also detected an interaction between the Snd2–Snd3 complex and the cytosolic (that is, non translocated and unglycosylated) fraction of our model substrate, RFP–Gas1 (Fig. 1d).

Next, we found that the stability of Snd1 protein was compromised in the *SND2/SND3* double mutant (Fig. 1e). Conversely, localization of Snd2 and Snd3 was altered upon loss of other SND components (Extended Data Fig. 2b). We observed a synthetic sick interaction between $\Delta snd3$ and $\Delta snd1$ or $\Delta snd2$ when inspecting colony sizes (Extended Data Fig. 2c), but complete epistasis of $\Delta snd3$ with both $\Delta snd1$ and $\Delta snd2$ in their effect on Gas1 aggregation (Extended Data Fig. 2d, e). As the single deletion of *SND3* led to impaired growth (Extended Data Fig. 2c), and as the Snd3 protein is an order of magnitude more abundant than Snd1 or Snd2 (ref. 24), it is possible that its synthetic growth interaction results from an additional cellular role, unrelated to its common function with *SND1* and *SND2*. Together, these findings support the idea that the SND proteins function in a joint targeting pathway.

¹Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel. ²Department of Cellular and Molecular Pharmacology, UCSF California Institute for Quantitative Biomedical Research and Howard Hughes Medical Institute, San Francisco, California 94158-2330, USA. ³Department of Molecular Biology, University Medical Center Göttingen, 37073 Göttingen, Germany. ⁴Department of Medical Biochemistry and Molecular Biology, Saarland University, 66421 Homburg, Germany. ⁵Max-Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany. [†]Present address: Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02114, USA.

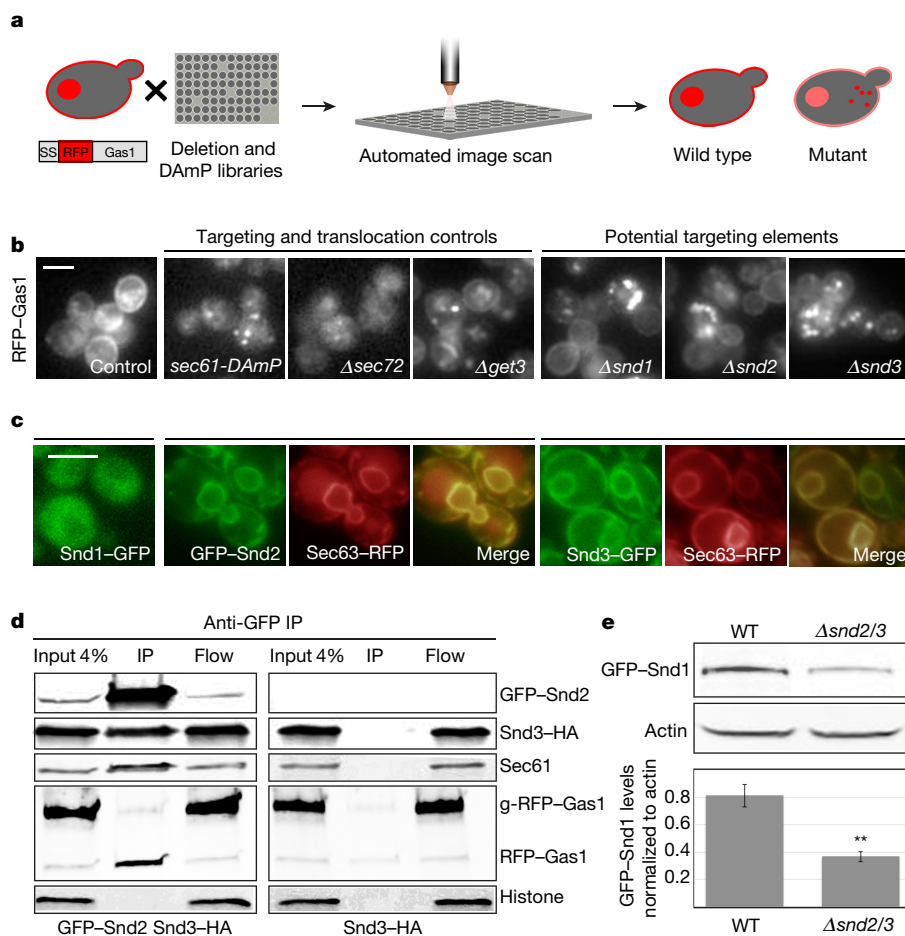


Figure 1 | A systematic screen uncovers uncharacterized ER targeting elements. **a**, A systematic screen for localization of SS-RFP-Gas1 on the background of yeast mutant libraries. SS, signal sequence. **b**, Mutants of *SND1*, *SND2* or *SND3* affect SS-RFP-Gas1 targeting similarly to known translocation or targeting mutants. **c**, Localization of GFP-tagged Snd1, Snd2 and Snd3. ER is marked by Sec63-RFP. **d**, Anti-GFP immunoprecipitation from GFP-Snd2/Snd3-HA strain and the negative

control Snd3-HA strain. GFP-Snd2 co-immunoprecipitated with Snd3-HA, Sec61 and the uninserted, cytosolic form of RFP-Gas1. **e**, GFP-Snd1 levels are lower in the Δ snd2/ Δ snd3 double mutant compared to the wild type (WT). Data shown are means \pm s.e.m. $^{**}P < 0.01$ by two-tailed Student's *t*-test, $n = 3$, biological replicates. For gel source images, see Supplementary Fig. 1. Scale bars, 5 μ m. All images are representative of around 300 cells per strain.

To directly test whether SND proteins affect targeting and uncover their substrate range, we turned to proximity-specific ribosome profiling²⁵ (Fig. 2a). A subset of transcripts was depleted on the ER membrane in cells from all three Δ snd strains, providing independent evidence that the SND machinery has a role in targeting these substrates to the ER surface as they are being translated (Fig. 2b). The proteins most affected in the *snd* mutants have been previously shown to accumulate on the ER membrane in the presence of the translational inhibitor cycloheximide, indicative of targeting that is only loosely coupled to translation²⁵. Notably, transcripts encoding proteins bearing an N-terminal transmembrane domain (TMD) (that is, in the first 95 amino acids, within the optimal recognition window of SRP²⁶) did not appear to be effected, whereas proteins with more downstream TMDs were depleted in all three Δ snd strains (Fig. 2c; for a full list of transcripts affected in *SND* mutants, see Supplementary Table 3).

To verify the results of the ribosome-profiling assay, we assayed a representative secretory protein, Ynl181w, whose TMD is in the central portion of the protein. Indeed, Ynl181w showed reduced targeting in the Δ snd strains by both microscopy and *in-vivo* translocation assays (Fig. 2d, e).

Our data suggest that the location of the first TMD within the protein is a major determinant of SND targeting and therefore merely altering the relative position of the first TMD in the protein should alter its pathway dependence. To test this idea, we re-engineered two secretory substrates to alter their TMD positions and tested their

targeting dependence. As predicted, simply moving the TMDs of two different substrates could alter their dependence on the SRP, SND or GET pathways (Fig. 2f, g and Extended Data Fig. 3).

Together, our results suggest either that SND proteins specifically recognize substrates with central TMDs or that SNDs cater for a broad substrate range, but their loss is only visible for those substrates not efficiently targeted by SRP or GET. To differentiate between these two possibilities, we investigated the relationship between the SNDs and the SRP and GET pathways.

To explore the interaction between the SNDs and the SRP pathway we used the previously described *sec65-1* temperature-sensitive strain, which has reduced SRP function at the permissive temperature and complete loss of SRP at the restrictive one. On this background we expressed the SND genes under the inducible *GAL1* promoter (Galp), which leads to either full repression (in glucose) or overexpression (in galactose). Slightly reduced SRP activity led to synthetic lethality in strains lacking *SND2* or *SND3* (Fig. 3a), whereas overexpression of *SND2* or *SND3* could rescue complete loss of SRP function (Fig. 3b), although Sec65 and Snd levels remained unchanged (Extended Data Fig. 4a, b). The restoration of viability was due to marked rescue of the cells' targeting capabilities by SND proteins (Fig. 3c, d and Extended Data Fig. 4c). These findings suggest that the SND proteins provide an alternate targeting route for a broad range of substrates, including targets that, under normal physiological conditions, are likely to be efficiently captured by SRP.

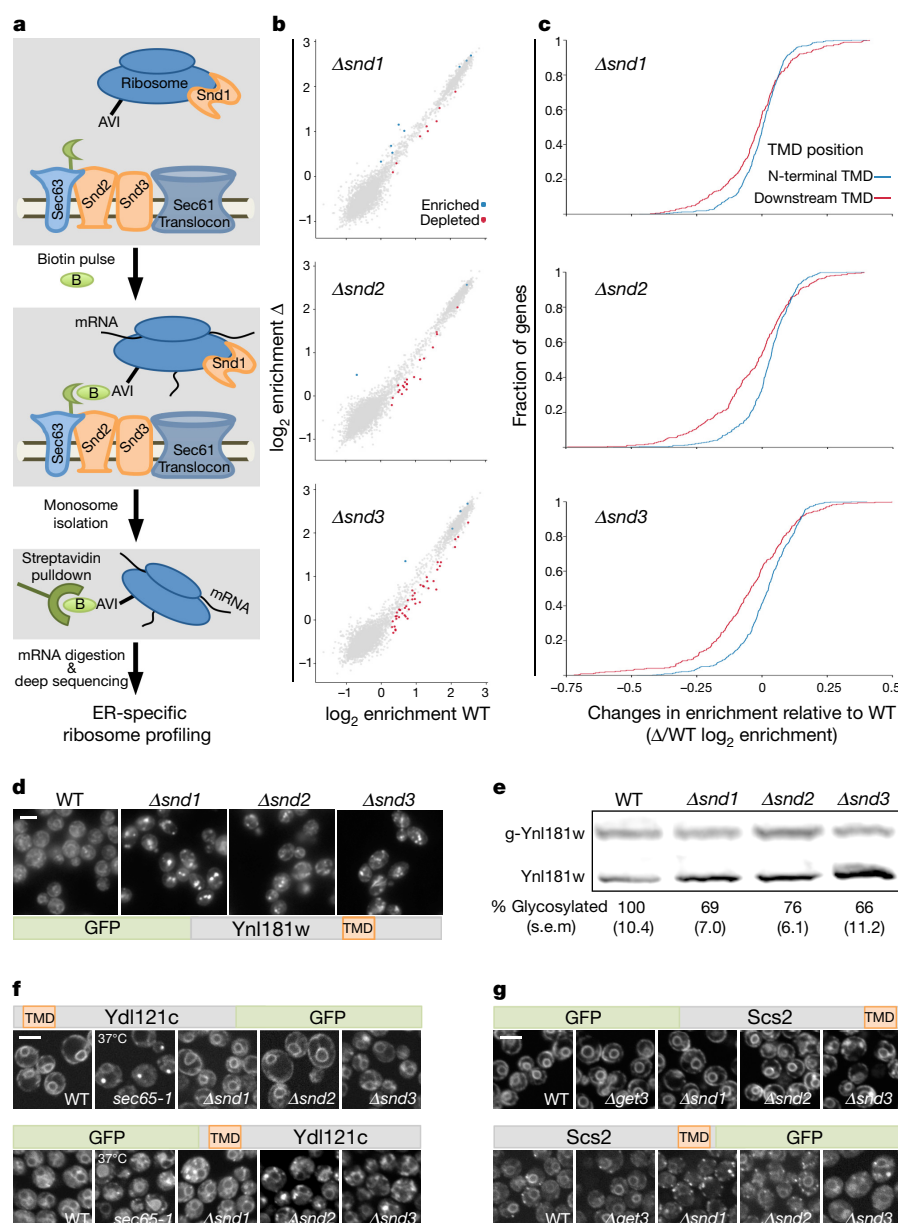


Figure 2 | SNDs affect the targeting of proteins with downstream transmembrane domains. **a**, Schematic of proximity-specific ribosome profiling. **b**, Translational enrichment on the ER surface. Significantly enriched or depleted transcripts in Δsnd mutants compared to wild type shown as blue or red circles, respectively. **c**, Cumulative distribution of ER enrichment of proteins with downstream TMDs (after 95 amino acids, red) or with an N-terminal TMD (in the first 95 amino acids,

If SND proteins act as a targeting pathway with a broad substrate range, then the SND and GET pathways should also act as functional backups for each other. As previously shown²⁷, deletion of *SND* genes becomes lethal in the absence of members of the GET pathway (Fig. 4a). However, concomitant loss of *SND* genes in strains lacking the auxiliary translocon (Δsec72) has epistatic or additive effects, suggesting that the SNDs target proteins to the SRP-independent translocon (Extended Data Fig. 5a).

To ascertain that the cause of death of double *SND*/GET mutants lay in drastic alterations of targeting efficiency, we created a conditional double mutant for *Snd2* and *Get3*, double deletion of which is synthetic lethal (Extended Data Fig. 5a, b). Metabolic labelling of RFP-Gas1 clearly demonstrated that although the single mutations (*Tetp-snd2* or Δget3) had only a small effect on translocation efficiency (Fig. 4b

blue). **d**, Microscopy images of GFP-Ynl181w localization in *SND* mutants. **e**, Western blot of translocation efficiency of glycosylatable HA (HA-Gly)-tagged Ynl181w (data shown are means (s.e.m.), $n = 3$, biological replicates). **g**, Ynl181w, glycosylated. **f**, **g**, Microscopy images of localization of re-engineered Ydl121c (**f**) or Scs2 (**g**) on the background of targeting mutants. For gel source images, see Supplementary Fig. 1. Scale bars, 5 μm . All images are representative of around 300 cells per strain.

and Extended Data Fig. 5c, d), the double mutant showed a marked decrease in mature RFP-Gas1 (Fig. 4b). The same is true for the known GET pathway substrate, the tail-anchored protein Ysy6 (ref. 6) (Fig. 4c), verifying our hypothesis that the two pathways work in parallel to ensure robust targeting. Conversely, when we used this system to test CPY or the SRP-dependent substrate DHC α F, their translocation was not hampered in either the single or the double mutants (Fig. 4d and Extended Data Fig. 5e–g), excluding any secondary effects of the double knockdown.

Supporting our hypothesis that SND and GET have functional redundancy for targeting proteins with downstream targeting signals, we find that a synthetic construct of GFP fused to the C-terminal glycosylphosphatidylinositol (GPI)-anchoring sequence of Gas1 (GFP-AS_{Gas1})⁹ can target well in the absence of either SND or GET but not in the

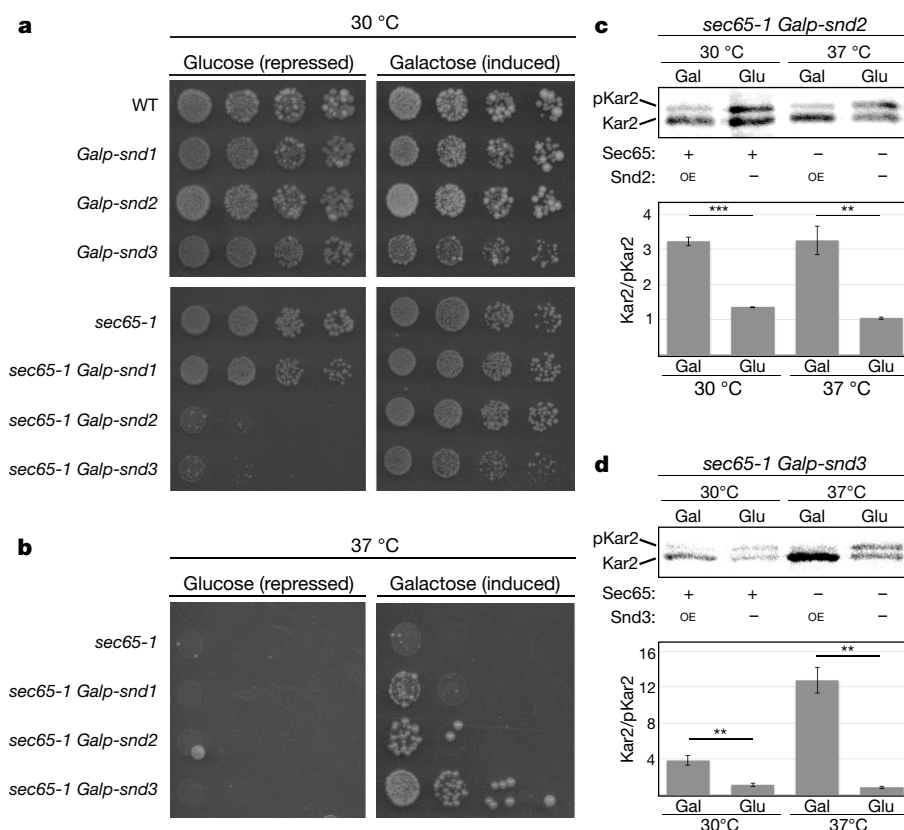


Figure 3 | SND proteins can compensate for loss of SRP. *SND* genes were expressed under the repressible (glucose) or inducible (galactose) Gal1 promoter. **a**, Growth at permissive temperature (30 °C, induces mild reduction in SRP). Repression of *SND*s leads to a synthetic sick or lethal phenotype. **b**, Growth at restrictive temperature (37 °C). Overexpression of *SND*s rescues lethality. **c**, **d**, Metabolic labelling of Kar2. When

either *SND2* (**c**) or *SND3* (**d**) was overexpressed, Kar2 was translocated significantly better than when *SND2* or *SND3* were repressed (data shown are means \pm s.e.m., $^{**}P < 0.01$, $^{***}P < 0.001$ by two-tailed Student's *t*-test, $n = 3$, biological replicates). Stronger *Snd3*-dependent translocation may explain the stronger rescue of this strain (**b**). For gel source images, see Supplementary Fig. 1.

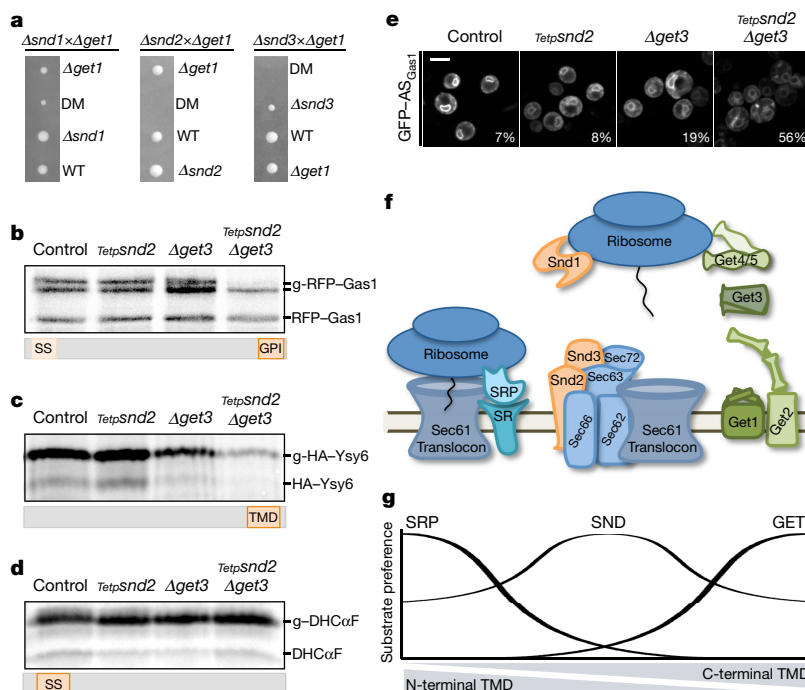


Figure 4 | The GET and SND pathways act as backups for targeting in vivo. **a**, Tetrads from $\Delta snd \Delta get$ diploids demonstrate a synthetic sick/lethal interaction. DM, double $\Delta snd \Delta get$ mutant. **b–d**, Metabolic labelling of RFP-Gas1 (**b**), HA-Ysy6 (**c**) and DHC α F (**d**) showing a decrease in translocated forms only for SRP-independent substrates in the conditional *SND2/GET3* double mutant. Accumulation of pre-inserted forms cannot

be observed owing to lack of proteasomal inhibition. Results reproduced in three biological replicates. **e**, GFP fused to Gas1 GPI-anchoring sequence (GFP-AS_{Gas1}). Percentage of cells (from 100) with mistargeting depicted on images. Scale bar, 5 μ m. **f**, Schematic of the eukaryotic ER-targeting apparatus. **g**, Model of the interplay among the ER-targeting pathways. For gel source images see Supplementary Fig. 1.

SND/GET double mutant (Fig. 4e). Hence, our results uncover the parallel functions of the SND and GET pathways in targeting proteins with downstream hydrophobic motifs.

Together, our data reveal that the SNDs can target diverse proteins with signals dispersed throughout the entire protein range. Moreover, they highlight a complex interconnected web with the two known targeting pathways that may explain the original difficulty in identifying the SND pathway.

Although its mechanistic details have yet to be uncovered, it is tempting to speculate about the mode of function of the SND pathway (Fig. 4f). As Snd1 is predicted to interact with ribosomes²⁰, it may be involved in capturing nascent substrates. Snd2 and Snd3, both ER transmembrane proteins that form a complex together with the translocon (Fig. 1d and Extended Data Fig. 2a), could act as receptors in promoting substrate capture and handoff to the translocon. It would be interesting to investigate whether the human orthologue hSnd2 has a similar role in mammalian cells.

Our data show that the three pathways, SRP, SND and GET, work in parallel to facilitate ER targeting of proteins bearing targeting motifs in all possible positions (Fig. 4g). If there is no distinct transmembrane position at which one pathway ceases to function and another steps in, but rather there exists a gradual decrease or increase in affinities, this would explain the substrate overlap between the pathways and the difficulties in reaching coherent understanding of the targeting requirements for many substrates⁷. More generally, identifying the role of the SND proteins demonstrates once again the intricacy and complexity of the cellular network engaged in efficient targeting.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 March; accepted 18 October 2016.

1. Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663–669 (2007).
2. Walter, P. & Johnson, A. E. Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10**, 87–119 (1994).
3. Favaloro, V., Spasic, M., Schwappach, B. & Dobberstein, B. Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. *J. Cell Sci.* **121**, 1832–1840 (2008).
4. Stefanovic, S. & Hegde, R. S. Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell* **128**, 1147–1159 (2007).
5. Jonikas, M. C. *et al.* Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–1697 (2009).
6. Schuldiner, M. *et al.* The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* **134**, 634–645 (2008).
7. Aviram, N. & Schuldiner, M. Embracing the void—how much do we really know about targeting and translocation to the endoplasmic reticulum? *Curr. Opin. Cell Biol.* **29**, 8–17 (2014).
8. Ast, T. & Schuldiner, M. All roads lead to Rome (but some may be harder to travel): SRP-independent translocation into the endoplasmic reticulum. *Crit. Rev. Biochem. Mol. Biol.* **48**, 273–288 (2013).
9. Ast, T., Cohen, G. & Schuldiner, M. A network of cytosolic factors targets SRP-independent proteins to the endoplasmic reticulum. *Cell* **152**, 1134–1145 (2013).
10. Ng, D. T., Brown, J. D. & Walter, P. Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J. Cell Biol.* **134**, 269–278 (1996).
11. Cohen, Y. & Schuldiner, M. Advanced methods for high-throughput microscopy screening of genetically modified yeast libraries. *Methods Mol. Biol.* **781**, 127–159 (2011).

12. Tong, A. H. Y. & Boone, C. Synthetic genetic array analysis in *Saccharomyces cerevisiae*. *Methods Mol. Biol.* **313**, 171–192 (2006).
13. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
14. Breslow, D. K. *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods* **5**, 711–718 (2008).
15. Breker, M., Gymrek, M. & Schuldiner, M. A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. *J. Cell Biol.* **200**, 839–850 (2013).
16. Mandon, E. C., Trueman, S. F. & Gilmore, R. Protein translocation across the rough endoplasmic reticulum. *Cold Spring Harb. Perspect. Biol.* **5**, 1–14 (2013).
17. Harada, Y., Li, H., Wall, J. S., Li, H. & Lennarz, W. J. Structural studies and the assembly of the heptameric post-translational translocon complex. *J. Biol. Chem.* **286**, 2956–2965 (2011).
18. Kaganovich, D., Kopito, R. & Frydman, J. Misfolded proteins partition between two distinct quality control compartments. *Nature* **454**, 1088–1095 (2008).
19. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
20. Fleischer, T. C., Weaver, C. M., McAfee, K. J., Jennings, J. L. & Link, A. J. Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes Dev.* **20**, 1294–1307 (2006).
21. Ricarte, F. *et al.* A genome-wide immunodetection screen in *S. cerevisiae* uncovers novel genes involved in lysosomal vacuole function and morphology. *PLoS One* **6**, e23696 (2011).
22. Zhao, Y. *et al.* Transmembrane protein 208: a novel ER-localized protein that regulates autophagy and ER stress. *PLoS One* **8**, e64228 (2013).
23. Yompakdee, C., Ogawa, N., Harashima, S. & Oshima, Y. A putative membrane protein, Pho88p, involved in inorganic phosphate transport in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **251**, 580–590 (1996).
24. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 (2014).
25. Jan, C. H., Williams, C. C. & Weissman, J. S. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **80**, 1257521 (2014). doi:10.1126/science.1257521
26. Noriega, T. R. *et al.* Signal recognition particle-ribosome binding is sensitive to nascent chain length. *J. Biol. Chem.* **289**, 19294–19305 (2014).
27. Pan, X. *et al.* A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell* **124**, 1069–1081 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Schuldiner laboratory members for discussions and comments on the manuscript; D. Kaganovich, T. Ravid, J. Gerst, S. High and H. Riezman for plasmids; P. Walter and M. Seedorf for antibodies; and I. Yofe and U. Weill for the N-terminal tagging plasmid and primers. T.A. was supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities. The work on human cells was supported by a DFG grant (IRTG 1830 and ZI 234/13-1) to R.Z. and the generation of anti-hSnd2 antibodies was funded by HOMFOR (HOMFOR2015). Supercomplex analysis by E.C.A. and B.S. was funded by the Deutsche Forschungsgemeinschaft (SFB 1190 P04). J.S.W. is supported by the NIH/NGMS (Center for RNA Systems Biology P50 GM102706 (Cate)). E.A.C. is supported by the National Science Foundation under grant 1144247. This work was funded by the Minerva foundation and Israel Science Foundation grant number 791/14 support to M.S. M.S. is an incumbent of the Dr. Gilbert Omenn and Martha Darling Professorial Chair in Molecular Genetics.

Author Contributions N.A., T.A. and M.S. conceptualized the study; S.G.C., E.A.C. and C.H.J. performed computational analysis; E.A.C. performed the ribosome profiling experiments, E.C.A. performed the blue-native page experiments; S.H., J.D., M.J. and S.S. performed the mammalian experiments; N.A. and T.A. performed all other experiments; T.A., B.S., R.Z., J.S.W. and M.S. supervised the study; N.A. and M.S. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. (maya.schuldiner@weizmann.ac.il) or T.A. (tast@broadinstitute.org).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Yeast strains and strain construction. Yeast strains are all based on the BY4741 laboratory strain²⁸. Manipulations were performed using a standard PEG/LiAC transformation protocol²⁹. All deletions were verified using primers from within the endogenous open reading frame. Primers for all genetic manipulations were planned either manually or using the Primers-4-Yeast web tool³⁰. All strains, primers and plasmids used in this study^{28,31–34} are listed in Supplementary Table 4.

Automated yeast library manipulations and high-throughput microscopy. SGA and microscopic screening were performed using an automated microscopy set-up as previously described^{11,15}, using the RoToR bench-top colony arrayer (Singer Instruments) and automated inverted fluorescent microscopic ScanR system (Olympus). Images were acquired using a 60× air lens with excitation at 490/20 nm and emission at 535/50 nm (GFP) or excitation at 575/35 nm and emission at 632/60 nm (RFP). After acquisition, images were manually reviewed using the ScanR analysis program.

Manual microscopy. Manual microscopy was performed using by one of two apparatuses. (I) Olympus IX71 microscope controlled by the Delta Vision SoftWoRx 3.5.1 software. Images were acquired using a 60× oil lens and captured by PhoeometricsCoolSnap HQ camera with excitation at 490/20 nm and emission at 528/38 nm (GFP/YFP) or excitation at 555/28 nm and emission at 617/73 nm (mCherry/RFP). (II) VisiScope Confocal Cell Explorer system, composed of a Zeiss Yokogawa spinning disk scanning unit (CSU-W1) coupled with an inverted Olympus IX83 microscope. Images were acquired using a 60× oil lens and captured by a connected PCO-Edge sCMOS camera, controlled by VisView software, with wavelength of 488 nm (GFP) or 561 nm (mCherry/RFP). Images were transferred to Adobe Photoshop CS6 for slight adjustments to contrast and brightness.

Co-immunoprecipitation. Lysates for immunoprecipitation were prepared from indicated strains in mid-logarithmic growth grown in YPD reach medium. Cells were harvested, washed in distilled water, and resuspended in lysis buffer (50 mM Tris HCl pH 7, 150 mM NaCl) supplemented with protease inhibitors (complete EDTA-free cocktail; Roche) and frozen in a drop-by-drop fashion in liquid nitrogen. Frozen cells were then pulverized in a ball mill (1 min at 30 Hz; Retsch) and thawed with nutation. Samples were thawed in 1 ml lysis buffer supplemented with protease inhibitors and 1% CHAPS (Sigma Aldrich) at 4 °C for 1 h. All samples were then clarified by centrifugation at 14,000g at 4 °C for 15 min. The remaining supernatant was added to GFP-trap (Chromotek) for 1 h followed by centrifugation at 1,000g at 4 °C for 3 min, and the supernatant was set aside as the flow through. Beads were washed three times with lysis buffer supplemented with protease inhibitors, and bound proteins were released from the beads by a 5-min incubation at 95 °C in sample buffer. The total protein lysate, the flow through and the immunoprecipitation (IP) fraction were analysed by western blotting.

Western blot analysis. Yeast proteins were extracted by either NaOH or TCA protocol as previously described^{9,35} and resolved on polyacrylamide gels, transferred to nitrocellulose membrane blots, and probed with primary rabbit/mouse antibodies against HA (BioLegend, 901502), GFP (Abcam ab290), RFP (Abcam ab62341), histone H3 (Abcam ab1791), actin (Abcam ab8224), Sec65 (kindly provided by P. Walter) or Sec61 (kindly provided by M. Seedorf). The membranes were then probed with a secondary goat-anti-rabbit/mouse antibody conjugated to IRDye800 or to IRDye680 (LI-COR Biosciences). Membranes were scanned for infrared signal using the Odyssey Imaging System. Images were transferred to Adobe Photoshop CS6 for slight adjustments to contrast and brightness.

Metabolic labelling. Newly synthesized yeast proteins were radioactively labelled *in vivo* by a 7–10 min pulse with [³⁵S]methionine at either 30 °C or 37 °C. Labelling was stopped by adding to the cells ice-cold TCA to a final concentration of 10%. Cells were then lysed and proteins were immunoprecipitated as previously described³⁶ with antibodies against RFP (Abcam, ab62341), HA (BioLegend, 901502), Kar2 (kindly provided by P. Walter) or CPY (Abcam, ab113685). Protease inhibitors (complete EDTA-free cocktail; Roche) were used throughout the extraction and immunoprecipitation process. Immunoprecipitated samples were resolved on polyacrylamide gels, which were then exposed to Phosphor Screen (GE Life Sciences) and scanned by phosphorimager. Translocation efficiency was calculated as $\left(\frac{\text{ER form}}{\text{total protein}} \right)_{\text{mutant}} / \left(\frac{\text{ER form}}{\text{total protein}} \right)_{\text{WT}}$. The statistical significance of

differences was measured using two-tailed student *t*-test with unequal variance, as indicated in the figure legends. For the Tetp-repression experiments, doxycycline (Sigma-Aldrich) was added to the overnight culture and to the back-dilution medium at a final concentration of 15 µg/ml.

Proximity-specific ribosome profiling. The ribosomal subunits RPL16a/b were conjugated to AVI-tag (biotin acceptor peptide), and Sec63 was conjugated to

BirA (biotin ligase), allowing the specific biotinylation and streptavidin pull-down of ribosomes in close physical proximity to the ER membrane. By comparing the ribosomal footprints obtained from the total ribosome fraction and the streptavidin-pulled fraction, we measured ER-localized translation enrichment. Biotin induction was carried out at mid-logarithmic growth phase in the presence of cycloheximide, which was added to the medium 2 min before the addition of biotin, at a final concentration of 100 µg/ml. To induce biotinylation, D-biotin was added to the medium to a final concentration of 10 nM and biotinylation was allowed to proceed for 2 min at the same temperature as growth. Cells were harvested by filtration onto 0.45 µm pore size nitrocellulose filters (Whatman), scraped from the membrane, and immediately submerged in liquid nitrogen. The following steps of monosome isolation, streptavidin pulldown of biotinylated ribosomes, and library generation were done as previously described²⁵.

Ribosome profiling computational analysis. *Footprint sequence.* Sequencing reads were demultiplexed and stripped of 3' cloning adapters using in-house scripts. Reads were mapped sequentially to Bowtie indices composed of rRNAs, tRNAs, and finally all chromosomes using Bowtie 1.1.0. Only uniquely-mapped, zero-mismatch reads from the final genomic alignment were used for subsequent analyses. These alignments were assigned a specific P-site nucleotide using a 15-nt offset from the 3' end of reads.

Gene enrichments. Gene-level enrichments were computed by taking the log₂ ratio of biotinylated footprint density (reads per million) within a gene coding sequence (CDS) over the corresponding density of matched input ribosome-profiling experiment. Yeast genes were excluded from all analysis if they met any of the following criteria: had fewer than 100 CDS-mapping footprints in the input sample of a particular experiment; were annotated as 'dubious' in the SGD database; mapped to the mitochondrial chromosome. Additionally, regions in which a CDS overlapped another same-strand CDS were excluded from enrichment calculations. *TMD classifications.* TMD positions were predicted using the Phobius algorithm. TMD classification was divided based on the start site of the first predicted TMD: N-terminal TMDs start in the first 95 amino acids of the protein; downstream TMDs start after the first 95 amino acids of the protein.

SND-dependent genes. Genes that were dependent on SND components were identified by comparing the Sec63-BirA ER enrichment in a wild-type strain (yJW1784) with that in a Δ snd strain (yJW1811, yJW1812, or yJW1813) as previously described²⁵. Briefly, log₂ enrichments were separately normalized by subtracting the mean enrichment and dividing by the standard deviation of enrichments for the corresponding experiment. Genes were then binned by the minimum number of sequencing counts in either wild-type or Δ snd input sample, and the difference between normalized enrichments was compared within each bin. Enriched genes were defined as those genes whose Δ snd log₂ enrichments were greater than 0.3 and whose enrichments increased in the Δ snd sample by at least two standard deviations compared to other genes in that bin. Depleted genes were defined as those genes whose wild type log₂ enrichments were greater than 0.3 and whose enrichments decreased in the Δ snd sample by at least two standard deviations compared to other genes in that bin. Significant depletion of 10–23%, 9–42% and 14–45% was observed in Δ snd1, Δ snd2 and Δ snd3, respectively. Including or excluding SS-bearing proteins had no effect on this trend. Mitochondrial proteins were excluded from the analysis.

Immunoprecipitation followed by mass spectrometry. Lysates for immunoprecipitations were prepared from yeast that expressed GFP-tagged SND genes or a constitutively expressed GFP-negative control, in mid-logarithmic growth grown in YPD reach medium. Cells were harvested, washed in distilled water, and resuspended in lysis buffer (50 mM Tris HCl pH 7, 150 mM NaCl) supplemented with protease inhibitors (complete EDTA-free cocktail; Roche) and frozen in a drop-by-drop fashion in liquid nitrogen. Frozen cells were then pulverized in a ball mill (1 min at 30 Hz; Retsch) and thawed with nutation. Samples were thawed in 1 ml lysis buffer supplemented with protease inhibitors and 1% digitonin (Sigma Aldrich) at 4 °C for 1 h. All samples were then clarified by centrifugation at 14,000g at 4 °C for 15 min. The remaining supernatant was added to GFP-trap (Chromotek) for 1 h followed by three washes with lysis buffer supplemented with protease inhibitors and 1% digitonin. Bound proteins were released from the beads by a 5-min acidic treatment (0.2 M glycine pH 2.5), which was neutralized with 1 M Tris pH 9.4. The eluted proteins were digested with 0.4 µg sequencing grade trypsin for 2 h in the presence of 100 µl of 2 M urea, 50 mM Tris HCl pH 7.5 and 1 mM DTT. The resulting peptides were acidified with trifluoroacetic acid (TFA) and purified on C18 StageTips. LC-MS/MS analysis was performed on an EASY-nLC1000 UHPLC (Thermo Scientific) coupled to a Q-Exactive mass spectrometer (Thermo Scientific). Peptides were loaded onto the column with buffer A (0.5% acetic acid) and separated on a 50-cm PepMap column (75 µm i.d., 2 µm beads; Dionex) using a 4-h gradient of 5–30% buffer B (80% acetonitrile, 0.5% acetic acid). Interactors were extracted by comparing the protein intensities to a GFP control.

Blue native gel electrophoresis. Yeast microsomes were extracted from the ADHp-SND2-GFP/SND3-HA strain as described³⁷. In brief, spheroplasts of yeast were lysed by dounce homogenization (25 strokes) in lysis buffer (0.1 M sorbitol, 20 mM HEPES pH 7.4, 50 mM potassium acetate, 2 mM EDTA, 1 mM DTT, 1 mM PMSF) at 4 °C. The lysates were centrifuged at 1,000g and the resulting supernatant at 27,000g for 10 min at 4 °C. The crude membrane pellet was re-suspended in lysis buffer and layered onto a discontinuous sucrose density gradient consisting of 1.2 and 1.5 M sucrose. Following centrifugation at 100,000g for 60 min at 4 °C, the membranes at the 1.2–1.5 M sucrose interface were collected and washed twice in lysis buffer. The membrane pellets were re-suspended in membrane storage buffer (50 mM NaCl, 0.32 M sucrose, 20 mM HEPES pH 7.4, 2 mM EDTA containing protease inhibitors) and the protein concentration determined by a standard Bradford assay.

Microsomes were solubilized in ComplexioLyte 48 buffer (1 mg/ml, Logopharm) for 30 min at 4 °C³⁸. Solubilized extracts were centrifuged at 100,000g for 30 min at 4 °C, supplemented with glycerol (5%) and coomassie G-250 (0.3%) and loaded on a 3.5–15% linear native polyacrylamide gel. The BN-PAGE gel was prepared as described³⁹. The gel buffer contained 25 mM imidazole and 500 mM 6-aminohexanoic acid. The cathode chamber was first filled with cathode buffer B (50 mM Tricine, 7.5 mM imidazole and 0.02% coomassie) and subsequently replaced by cathode buffer B/10 (containing 0.002% coomassie) after the gel running front had covered a third of the desired distance of electrophoresis. The anode chamber was filled with 25 mM imidazole pH 7.0. A high-molecular-weight calibration kit for native electrophoresis from GE Healthcare was used as a standard. For 2D BN-PAGE, the excised lanes were equilibrated in 2D-dissociation buffer (60 mM Tris/HCl pH 6.8, 10% glycerol, 2% SDS, 5% v/v β -mercaptoethanol, 6 M urea) before separation on the second dimension by SDS-PAGE. After electro-blotting, the nitrocellulose membrane was detected with the indicated antibodies.

Analysis of human cells. The HEK293 cell line used was obtained from DSMZ (no. ACC 305). DSMZ supplied verification of authentication of the cells, tested by DSMZ via short tandem repeat loci (STR profile). The cell line is routinely tested for mycoplasma contamination. This cell line was chosen as it is routinely used for fractionation experiments.

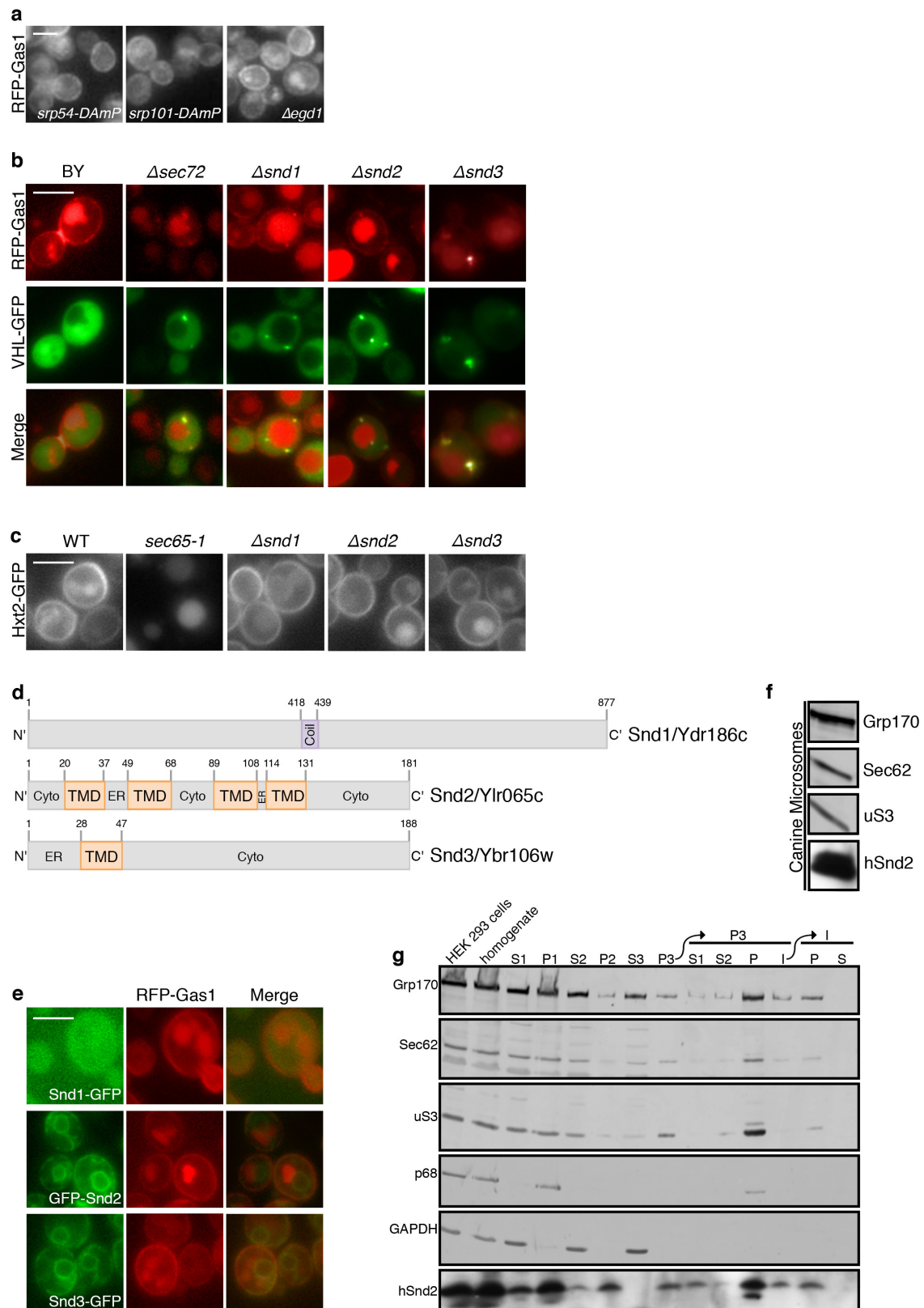
Rough microsomes from human cells were prepared as described⁴⁰. Briefly, 30×10^6 HEK293 cells were harvested and washed once with PBS and twice with buffer 1 (50 mM HEPES/KOH pH 7.5; 0.25 M sucrose; 50 mM KOAc; 6 mM MgOAc; 4 mM PMSF; 1 mM EDTA; 1 mM DTT; 0.1 mg/ml cycloheximide; 0.3 U/ml RNasin (Promega); protease inhibitor cocktail). After homogenization in buffer 1 using a glass/Teflon homogenizer, the suspension was centrifuged at 1,000g for 10 min. The supernatant was centrifuged at 10,000g for 10 min. The new supernatant was layered onto 0.6 M sucrose in buffer 2 (50 mM HEPES/KOH pH 7.5, 0.6 M sucrose, 100 mM KOAc, 5 mM MgOAc, 4 mM DTT, 0.1 mg/ml cycloheximide, 40 U/ml RNasin) and centrifuged at 230,000g for 90 min. The resulting membrane pellet was previously shown to comprise rough ER. Here, it was resuspended in buffer 2 and adjusted to 2.3 M sucrose, which was overlaid with 1.9 and 0 M sucrose, respectively, in buffer 2. After flotation at 100,000g for 18 h, the interphase between 0 and 1.9 M sucrose, two fractions of the remaining supernatant, and the pellet were collected. After centrifugation of the interphase

at 100,000g for 1 h, the membrane pellet corresponded to purified rough ER. All steps after the first washing step were carried out on ice.

Western blot analyses employed antibodies against β -actin (Sigma), CAML (Synaptic Systems SA7679), or rabbit antibodies that were raised against the depicted proteins: the C-terminal peptide of hSnd2 (KRQRRQERRQMKRL) plus an N-terminal cysteine; or an internal peptide of SR α (KKFEDSEKAKKPVVR) plus a C-terminal cysteine, cross-linked to KLH. The SR α and β -actin antibodies were visualized using ECL Plex goat-anti-rabbit IgG-Cy5-conjugate or ECL Plex goat-anti-mouse IgG-Cy3-conjugate (GE Healthcare) and the Typhoon-Trio imaging system (GE Healthcare) in combination with Image Quant TL software 7.0 (GE Healthcare). The hSnd2 and CAML antibodies were visualized using secondary peroxidase (POD)-coupled anti-rabbit antibody (Sigma) plus ECL (GE Healthcare) and the Fusion SL luminescence-imaging system (PqLab) in combination with Image Quant TL software 7.0.

Data availability. Ribosome-profiling data have been deposited in Gene Expression Omnibus (GEO) under accession number GSE85686. Gel source images can be found in Supplementary Fig. 1. Other data that support the findings of this study are available from the authors on reasonable request.

28. Brachmann, C. B. *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1998).
29. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87–96 (2002).
30. Yofe, I. & Schuldiner, M. Primers-4-Yeast: a comprehensive web tool for planning primers for *Saccharomyces cerevisiae*. *Yeast* **31**, 77–80 (2014).
31. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
32. Kitada, K., Yamaguchi, E. & Arisawa, M. Cloning of the *Candida glabrata* TRP1 and HIS3 genes, and construction of their disruptant strains by sequential integrative transformation. *Gene* **165**, 203–206 (1995).
33. Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* **15**, 1541–1553 (1999).
34. Stirling, C. J. & Hewitt, E. W. The *S. cerevisiae* SEC65 gene encodes a component of yeast signal recognition particle with homology to human SRP19. *Nature* **356**, 534–537 (1992).
35. Kushnirov, V. V. Rapid and reliable protein extraction from yeast. *Yeast* **16**, 857–860 (2000).
36. Graham, T. R. Metabolic labeling and immunoprecipitation of yeast proteins. *Curr. Protoc. Cell Biol.* **7**, 7.6 (2001).
37. Wuestehube, L. J. & Schekman, R. W. Reconstitution of transport from endoplasmic reticulum to Golgi complex using endoplasmic reticulum-enriched membrane fraction from yeast. *Methods Enzymol.* **219**, 124–136 (1992).
38. Schwenk, J. *et al.* Functional proteomics identify cornichon proteins as auxiliary subunits of AMPA receptors. *Science* **323**, 1313–1319 (2009).
39. Wittig, I., Braun, H.-P. & Schägger, H. Blue native PAGE. *Nat. Protocols* **1**, 418–428 (2006).
40. Pfeffer, S. *et al.* Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. *Nat. Commun.* **5**, 3072 (2014).

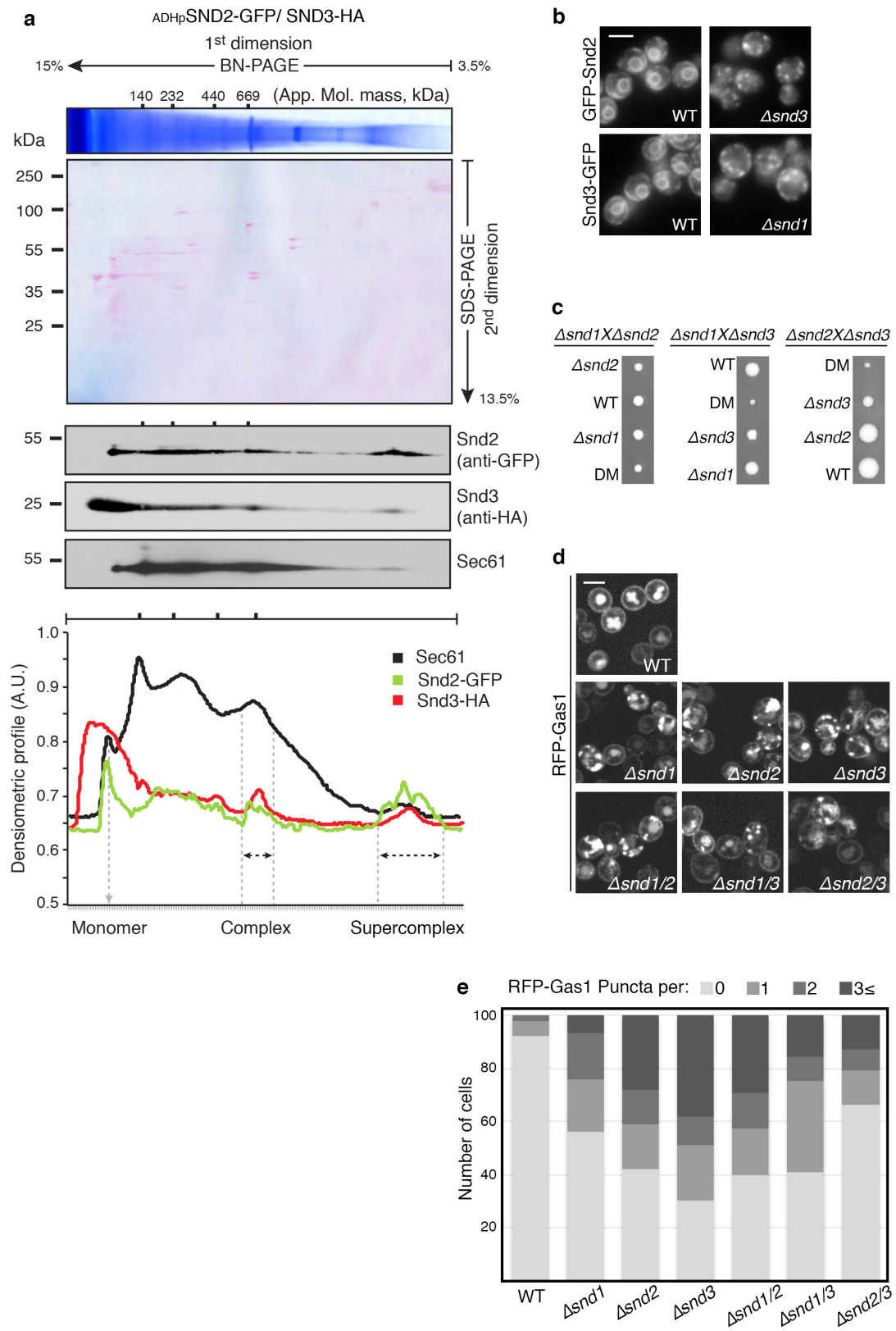


Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | The SND proteins mediate SRP-independent targeting.

a, Fluorescent micrographs of RFP–Gas1 confirm that it is not mislocalized when components of SRP, SRP receptor or NAC are compromised (control image can be found in Fig. 1b). Scale bars throughout figure, 5 μ m. Images throughout figure are representative of about 300 cells captured per strain. **b**, *SND* mutants accumulate RFP–Gas1 in inclusions. Fluorescent micrographs confirm that accumulations of RFP–Gas1 in Δ *snd* strains colocalize with the cytosolic inclusion marker VHL–GFP. **c**, *SND* deletions do not have a non-specific effect on translation, targeting or translocation. A fluorescently tagged SRP substrate (Hxt2–GFP) was mislocalized only in the temperature-sensitive strain *sec65-1* when grown at the restrictive temperature of 37 °C (under these conditions the cells lack functional SRP). *SND*-deleted strains display normal cell surface localization of Hxt2. **d**, Schematic representation of the structural elements and topology predictions of Snd1 (top), Snd2

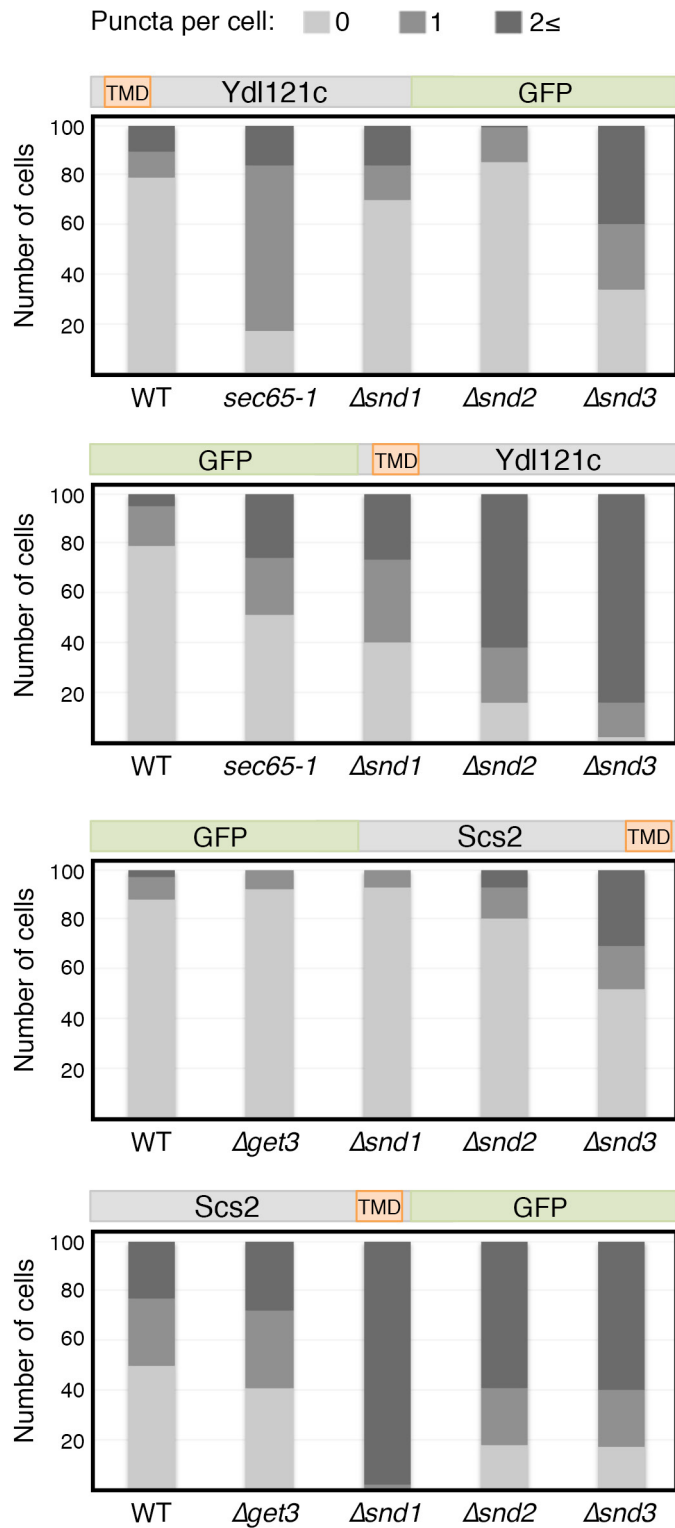
(middle) and Snd3 (bottom). Numbers indicate the number of amino acids in the proteins. **e**, RFP–Gas1 is correctly localized in all GFP-tagged *SND* proteins, indicating that the tag does not disrupt their function and endogenous localization. **f**, A mammalian orthologue of Snd2 (hSnd2) is present in canine pancreatic rough microsomes, which are routinely used as a source of mammalian ER proteins, as seen by immunoblotting with an antibody (data not shown) against hSnd2 that was shown to be specific by siRNA-mediated gene silencing. **g**, Endogenous hSnd2 is localized to human rough ER. HEK293 cells were homogenized and subfractionated into various pellet (P) and supernatant (S) fractions. Fractions were analysed by SDS–PAGE and immunoblotting. hSnd2 co-fractionated with the rough ER markers Grp170 and Sec62 and the ribosomal protein uS3 but not with the nuclear and cytosolic proteins p68 and GAPDH. The areas of interest of luminescence images from a single western blot are shown. For gel source images see Supplementary Fig. 1.



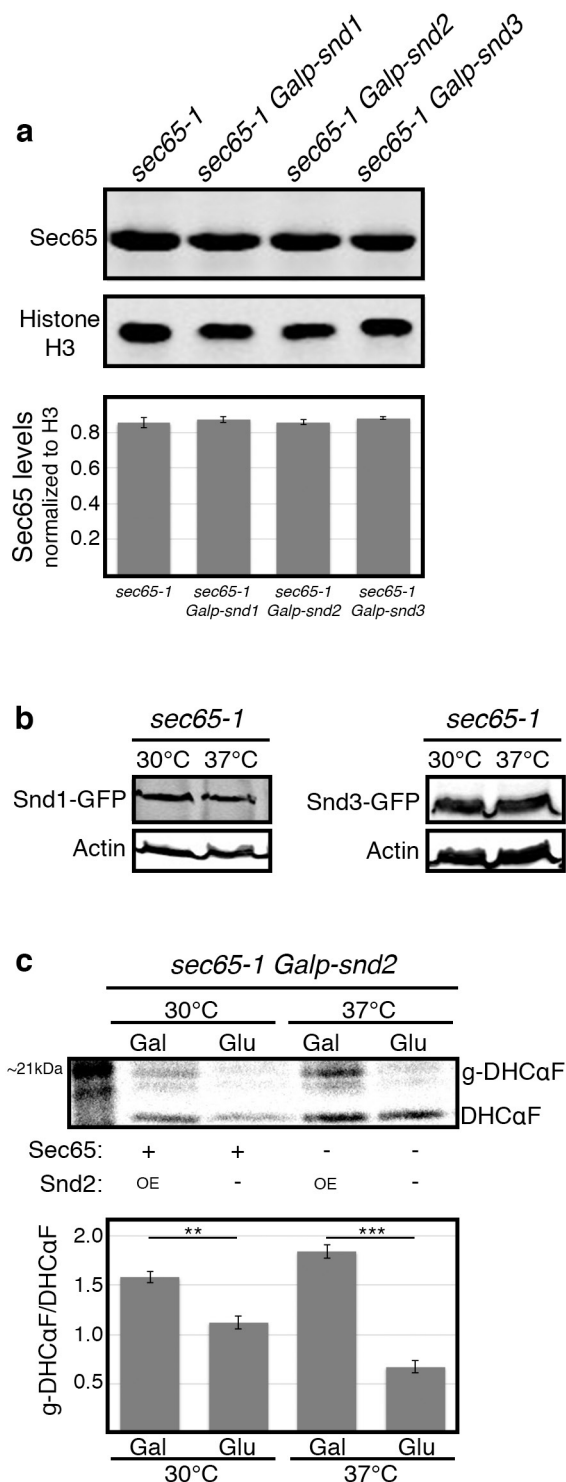
Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | The SND proteins function in a joint targeting pathway. **a**, Snd2 and Snd3 form a complex with the Sec61 translocon, as shown by BN-PAGE followed by second-dimension SDS-PAGE. Densitometry quantification revealed that Sec61 migrates in four distinct complexes as well as a monomer. Snd2 and Snd3 reside together in two of these complexes, one approximately 669 kDa and a second with a higher molecular mass. We postulate that the two Sec61-SND complexes may differ in size depending on the presence of additional auxiliary components. For gel source images see Supplementary Fig. 1. **b**, Fluorescent micrographs showing that Snd2 is mislocalized upon deletion of *SND3* and Snd3 is mislocalized upon deletion of *SND1*, suggesting functional dependence among the three proteins. Scale bars throughout figure, 5 μ m. Images throughout figure are representative

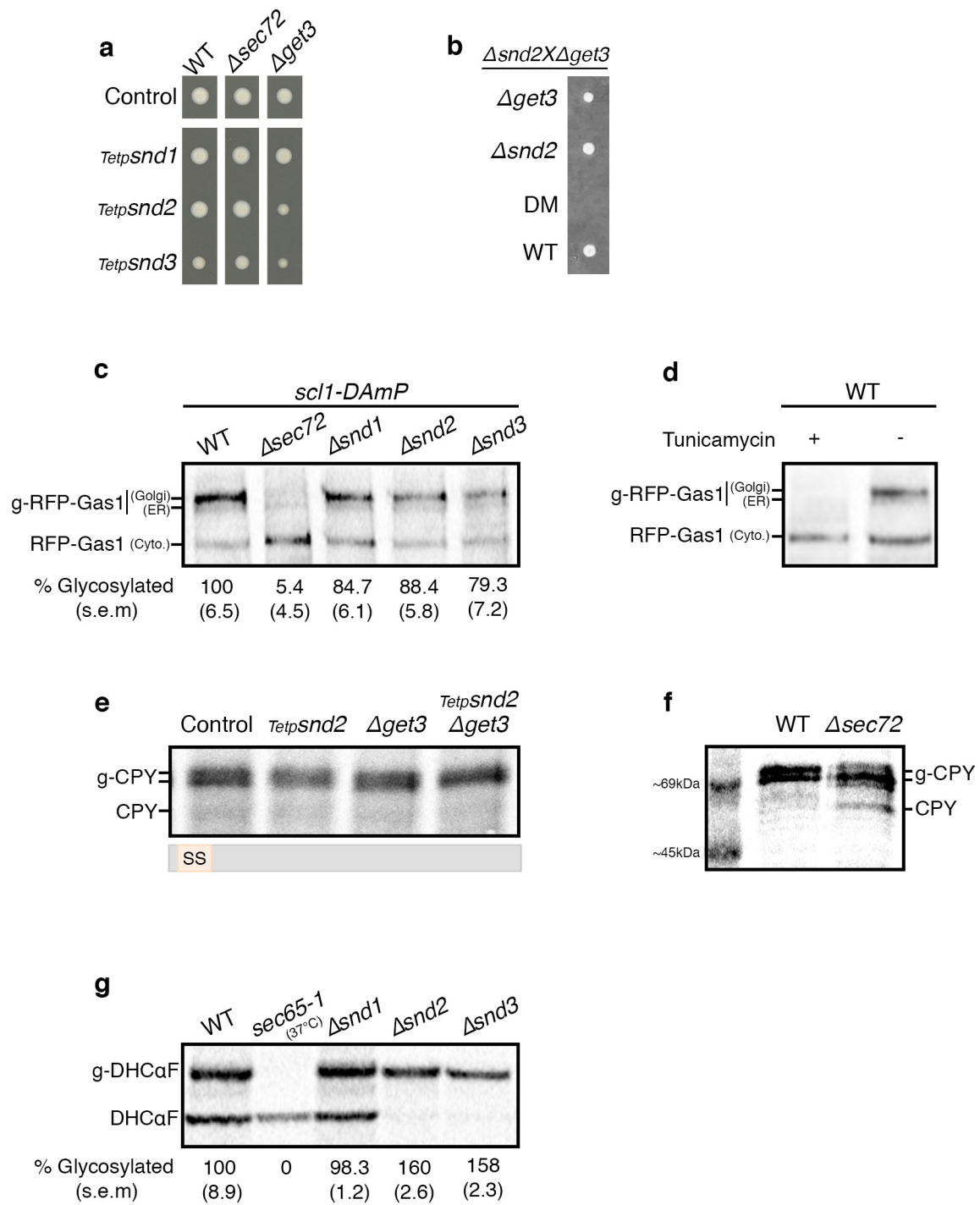
of around 300 cells captured per strain. **c**, Growth rates reveal genetic interactions among the *SND* genes. Heterozygous diploids of Δ *snd* were sporulated and tetrad-dissected to retrieve haploids. Tetrads obtained demonstrate an epistatic interaction between *SND1* and *SND2* mutations, and a synthetic sick interaction between *SND3* and the *SND1/2* mutations. As *SND3* is more than an order of magnitude more abundant than *SND1/2*, it is possible that this interaction is due to an independent cellular function. **d**, Fluorescent micrographs of RFP-Gas1 in single and double *SND* mutants show that the mutations are epistatic to each other in terms of their effect on targeting. **e**, Quantification of the RFP-Gas1 mislocalization phenotype in *SND* single and double mutants (Extended Data Fig. 2d) reveals a buffering epistatic interaction between *SND* genes (100 cells were counted per strain).



Extended Data Figure 3 | Substrate affinity to a targeting pathway depends on the position of its transmembrane domain. Quantification of the mislocalization phenotype in Fig. 2f, g confirms that re-positioning of a substrate's TMD can alter its dependence on the different targeting pathways.



Extended Data Figure 4 | Compensation for loss of SRP by the SNDs is not due to alteration in protein levels. **a**, Overexpression of *SND* genes does not affect SRP levels. *SND* genes were overexpressed by growth on galactose in 30 °C, and levels of Sec65 protein were measured by western blotting and normalized to a histone H3 loading control. No apparent change in *sec65-1* levels was detected, implying that the rescue observed in Fig. 3b–d is not due to increased SRP levels (data shown as mean \pm s.e.m., $n = 3$, no statistically significant difference was seen between the samples, biological replicates). **b**, Levels of *SND* proteins do not change in SRP-depleted cells. *SND* proteins were C-terminally tagged on the *sec65-1* background, and their levels were measured by western blotting when grown at either permissive or restrictive temperatures (30 °C and 37 °C, respectively), and normalized to an actin loading control. No apparent change in Snd1 or Snd3 levels was observed. Snd2 levels were below detection threshold (data not shown). **c**, Pulse radioactive metabolic labelling followed by DHcA immunoprecipitation was used to measure the translocation rate of DHcA. *SND2* overexpression induced significantly higher translocation when compared to its repression by glucose, regardless of the functional state of *sec65-1*. Data shown as mean \pm s.e.m. $**P < 0.01$, $***P < 0.001$, by two-tailed Student's *t*-test, $n = 3$, biological replicates. For all gel source images see Supplementary Fig. 1.



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | The GET and SND pathways are partially redundant. **a**, Repression of *SND* genes is epistatic with *SEC72* deletion and synthetic sick with *GET3* deletion. Growth rates of strains with the *SND* genes expressed under the regulation of a repressible Tet-promoter when grown on tetracycline. The growth rate of $\Delta sec72$ *Tetp-SNDs* conditional double mutants is identical to that of the control strain, indicating that these mutations are epistatic. The $\Delta get3$ *Tetp-SNDs* conditional double mutants are sick, but viable. **b**, Double deletion of *SND2* and *GET3* is lethal. Heterozygous diploids of $\Delta snd2$ and $\Delta get3$ were sporulated and tetrad-dissected to retrieve haploids. Tetrads obtained demonstrate a synthetic lethal interaction between *SND2* and *GET3*. **c**, RFP-Gas1 translocation is moderately affected by *SND* single deletions. Pulse radioactive metabolic labelling followed by RFP-Gas1 immunoprecipitation was used to measure RFP-Gas1 translocation rates. Percentage of glycosylated ER and Golgi forms (indicated by two black lines) was reduced to 5% in $\Delta sec72$ mutants, while in $\Delta snd1$, $\Delta snd2$ and $\Delta snd3$ strains it was reduced to 85%, 88% and 79%, respectively (data shown as mean (s.e.m.), $n = 3$, biological replicates). All strains in this assay were attenuated for degradation with the *scl1-DAmP* proteasome hypomorphic allele. **d**, Pulse radioactive metabolic labelling followed

by RFP-Gas1 immunoprecipitation was performed in the presence and absence of the glycosylation inhibitor tunicamycin, allowing the identification of three forms of RFP-Gas1: cytosolic, ER and Golgi (mature). **e**, CPY targeting is not affected by double mutants of the *SND* and GET pathways. The same methodology as in **c** was used to follow the signal sequence (SS)-containing protein CPY in the conditional *SND2/GET3* double mutant. A mild decrease in the glycosylated forms was observed in the *SND2* single mutant, but there was no translocation defect in the *GET3* single mutant or in the conditional double mutant. This result was repeated in three independent biological replicates. **f**, CPY was metabolically labelled in a control strain and a partially translocated pool was visualized with a ladder to provide a size reference to **e**. **g**, The same methodology as in **c** was used to measure the translocation rate of the SRP-dependent substrate DHCoF. In the temperature-sensitive strain *sec65-1*, at the restrictive temperature (37°C), there was no translocated substrate. The translocation efficiency of the $\Delta snd1$ strain was comparable to that of the wild-type control; in the $\Delta snd2$ and $\Delta snd3$ strains it was significantly higher (~160% glycosylated protein compared to the wild-type control). Data shown as mean (s.e.m.), $n = 3$, biological replicates. For all gel source images see Supplementary Fig. 1.

Multiple dynamin family members collaborate to drive mitochondrial division

Jason E. Lee¹, Laura M. Westrate¹, Haoxi Wu¹, Cynthia Page¹ & Gia K. Voeltz¹

Mitochondria cannot be generated *de novo*; they must grow, replicate their genome, and divide in order to be inherited by each daughter cell during mitosis. Mitochondrial division is a structural challenge that requires the substantial remodelling of membrane morphology^{1–3}. Although division factors differ across organisms, the need for multiple constriction steps and a dynamin-related protein (Drp1, Dnm1 in yeast) has been conserved^{4–6}. In mammalian cells, mitochondrial division has been shown to proceed with at least two sequential constriction steps: the endoplasmic reticulum and actin must first collaborate to generate constrictions suitable for Drp1 assembly on the mitochondrial outer membrane; Drp1 then further constricts membranes until mitochondrial fission occurs^{2,7–9}. *In vitro* experiments, however, indicate that Drp1 does not have the dynamic range to complete membrane fission⁷. In contrast to Drp1, the neuron-specific classical dynamin dynamin-1 (Dyn1) has been shown to assemble on narrower lipid profiles and facilitate spontaneous membrane fission upon GTP hydrolysis^{10,11}. Here we report that the ubiquitously expressed classical dynamin-2 (Dyn2) is a fundamental component of the mitochondrial division machinery. A combination of live-cell and electron microscopy in three different mammalian cell lines reveals that Dyn2 works in concert with Drp1 to orchestrate sequential constriction events that build up to division. Our work underscores the biophysical limitations of Drp1 and positions Dyn2, which has intrinsic membrane fission properties, at the final step of mitochondrial division.

Mitochondria are organelles that are ancestrally related to α -proteobacterium that have been adapted by eukaryotic cells to execute a broad range of cellular functions, including ATP production and apoptosis^{6,12}. The divergence of mitochondria in higher-order eukaryotes from those of primitive eukaryotes is marked by the loss of the FtsZ ring, a functional remnant of the bacterial cell-division machinery⁶. By contrast, the dynamin family of large GTPases, which regulates membrane dynamics in all eukaryotes, expanded to accommodate the growing demands of more complex organisms^{6,13}. For example, mitochondria in mammalian cells have a set of dynamin-related proteins that regulate mitochondrial dynamics: Drp1 for fission, and mitofusins (Mfn1/2) and optic atrophy-1 (Opa1) for fusion¹³. Despite an apparent drive towards the specialization of dynamins, the classical Dyn2 in mammalian cells (best known for its role in driving endocytosis at the plasma membrane) is more promiscuous, facilitating membrane fission events at multiple organelles and mediating the final stages of cytokinesis¹³. We therefore explored the possibility that Dyn2 could have a role in mitochondrial dynamics.

The depletion of Drp1 halts mitochondrial division and shifts dynamics towards fusion, resulting in an elongated, hyper-fused mitochondrial network¹⁴. To investigate whether similar results would be observed on Dyn2 depletion, monkey COS-7 and human HeLa cells were treated with small interfering RNAs (siRNAs) targeted to Dyn2 and compared to scrambled siRNA-treated (control) and Drp1 siRNA-treated cells (Fig. 1a and Extended Data Fig. 1, respectively).

Dyn2-depleted cells displayed a hyper-fused mitochondrial network that was similar to mitochondria in Drp1-depleted cells (Fig. 1a–c). Dyn2 depletion did not affect the steady-state levels of other fission and fusion factors, including Drp1, Mfn, Opa1 and Mfn2 (Extended Data Fig. 1d). To test whether mitochondrial elongation was a secondary effect of blocking clathrin-mediated endocytosis, we depleted the clathrin adaptor protein AP-2 (ref. 13). AP-2-depleted cells displayed normal mitochondrial morphology (Extended Data Fig. 2). Thus, the mitochondrial phenotype induced by Dyn2 depletion is not a consequence of inhibiting endocytosis. Alternatively, an endocytosis blockade could deplete cell nutrients. To test this, we measured whether Dyn2 depletion triggers starvation-induced phosphorylation of Drp1 at Ser637 of Drp1 (ref. 15). Dyn2 depletion slightly decreases phosphorylation at this inhibitory site, suggesting that it does not phenocopy the effects of starvation (Extended Data Fig. 3a). We did measure a reduction in mitosis-induced phosphorylation at the activating Ser616 of Drp1 (ref. 16); however, this could be due to a Dyn2-dependent mitotic block¹⁷ (Extended Data Fig. 3b). Overall, these data are consistent with a direct role for Dyn2 in mitochondrial fission.

There are structural differences between Dyn2 and Drp1 that translate into functional differences between the two proteins^{18,19}. Dyn2 has a Pleckstrin homology (PH) domain that is important for membrane insertion and a proline-rich domain (PRD) that is responsible for localization (Fig. 1d). Drp1 lacks both of these domains, instead requiring a variable domain to interact with adaptors (such as Mfn) that are responsible for recruiting Drp1 to mitochondria¹⁹. To gain mechanistic insight into the relationship between Dyn2 and mitochondria, we tested wild-type Dyn2, the GTP-binding-defective Dyn2(K44A), the PH-domain-deleted Dyn2(Δ PH), and the PRD-deleted Dyn2(Δ PRD) variants for their ability to rescue mitochondrial morphology following Dyn2 depletion. Simultaneous depletion of endogenous Dyn2 with the unabated expression of exogenous siRNA-resistant Dyn2–mNeon fusion proteins was verified by immunoblot analysis (Fig. 1e). Control cells, expressing both endogenous Dyn2 and exogenous wild-type Dyn2–mNeon, displayed normal mitochondrial morphology. By contrast, Dyn2–siRNA-treated cells expressing mNeon alone were depleted of Dyn2, predominantly displaying long mitochondria. Expression of exogenous wild-type Dyn2, but not Dyn2(K44A), Dyn2(Δ PH) or Dyn2(Δ PRD), was sufficient to reverse the elongated mitochondria phenotype that was induced by depletion of Dyn2. (Fig. 1f, g). Thus, the classical Dyn2 features GTP binding, recruitment, and lipid association, which are essential for Dyn2 functions at other organelles¹³, are also required for mitochondrial division.

We next performed live-cell imaging of Dyn2 to observe whether Dyn2 is recruited to Drp1-marked mitochondrial constrictions destined for division, since classical Dyn1 assembles on narrow tubules to drive their constriction and vesiculation *in vitro*^{10,11}. Dyn2 was observed at Drp1-marked mitochondrial division events in human skin melanoma cells (Sk-Mel2) genetically edited to express Dyn2–GFP under its endogenous promoter, and in *Cercopithecus aethiops* COS-7 and *Potorous tridactylus* PtK1 cells overexpressing Dyn2–mNeon

¹Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder, Boulder, Colorado 80309, USA.

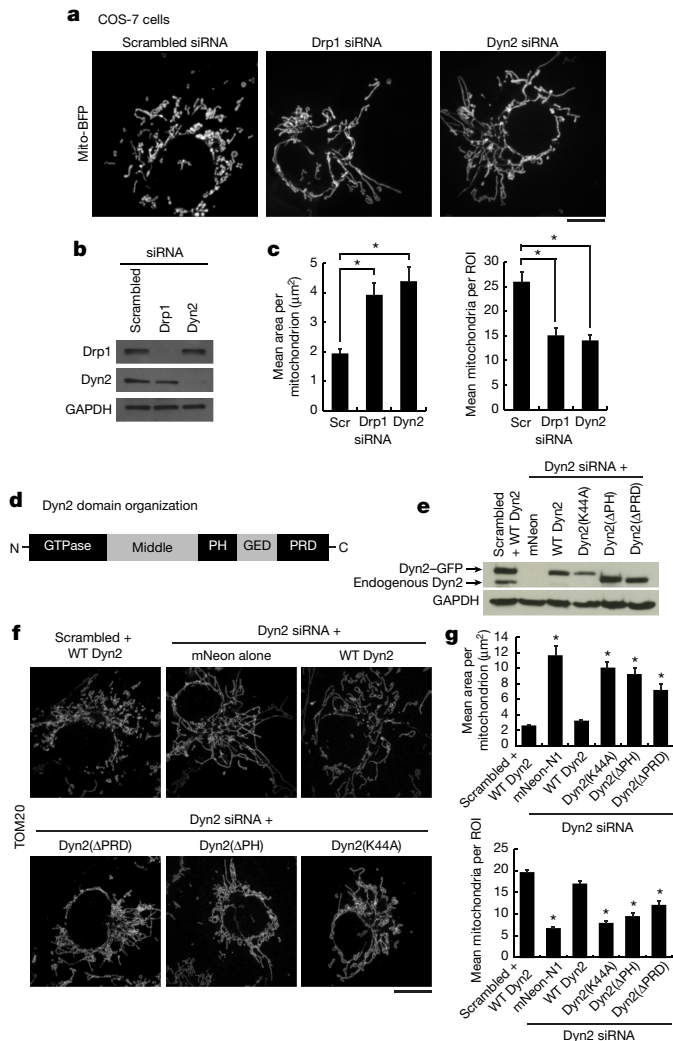


Figure 1 | Dyn2 is required for mitochondrial division. **a**, Representative images of blue fluorescent protein-labelled mitochondria (mito-BFP) COS-7 cells treated with scrambled siRNA ($n = 42$ cells), Drp1 siRNA ($n = 42$) and Dyn2 siRNA ($n = 43$). **b**, Immunoblot of Drp1, Dyn2, and GAPDH in cells treated with the indicated siRNA. **c**, Mitochondrial morphology was quantified for mean mitochondrial area per mitochondrion and mean mitochondria per region of interest (ROI). **d**, The domain organization of Dyn2. **e**, Immunoblot of endogenous and exogenous dynamin-expression levels in rescue experiments. GED, dynamin GTPase effector domain. **f**, Representative images of mitochondrial import receptor subunit TOM20 immunofluorescence in cells transfected with scrambled siRNA and wild-type (WT) Dyn2-mNeon ($n = 40$ cells), Dyn2 siRNA and N1-mNeon ($n = 40$), Dyn2 siRNA and WT Dyn2 ($n = 45$), Dyn2 siRNA and Dyn2(K44A)-mNeon ($n = 40$), Dyn2 siRNA and Dyn2(ΔPH)-mNeon ($n = 44$), Dyn2 siRNA Dyn2(ΔPRD)-mNeon ($n = 41$). **g**, Mitochondrial morphology was assessed as in **c**. Error bars represent s.e.m.; * $P < 0.01$ statistical significance calculated by ANOVA and obtained from three biological replicates. Scale bars, $10 \mu\text{m}$ (**a**, **f**).

(Fig. 2a, c, Extended Data Fig. 4a and Supplementary Videos 1–3). PtK1 cells are advantageous here as they lie flatter than Sk-Mel2 cells and their mitochondria divide more frequently than those in COS-7 cells. Thus, PtK1 cells were used to characterize the dynamics of Dyn2 and the mitochondrial division factors Mff and Drp1 during division (Fig. 2c–g and Extended Data Fig. 4d–f). Dyn2 puncta were observed at 80.8% of Drp1-marked mitochondrial division events ($n = 26$ events from 254 cells; Fig. 2f). Additionally, Drp1 puncta were present at the division site for a longer period of time than Dyn2 puncta, both before (Drp1, 52.33 ± 7.14 s; Dyn2, 27 ± 5.07 s) and after (Drp1, 56.67 ± 7.83 s; Dyn2, 31.67 ± 4.52 s) fission (Fig. 2g). The pre-fission lifetimes of Dyn2

puncta at mitochondria (Fig. 2g) are similar to those reported for Dyn2 during clathrin-mediated endocytosis at the cell surface in genome-edited cell lines (16–32 s) (ref. 20), suggesting that similar mechanisms of Dyn2-mediated membrane scission occur at both organelles.

A comparison of Drp1 and Dyn2 dynamics leading up to, during and following mitochondrial division uncovers differences in behaviour between the two dynamins. Drp1 puncta accumulate at mitochondrial constrictions regardless of whether they are about to undergo fission or not, whereas Dyn2 puncta are rarely found on mitochondria and appear to be recruited transiently, just before division. Upon division, Drp1 puncta split to maintain association with each newly created daughter mitochondrion. As would be expected for a Drp1 adaptor, Mff puncta mirror the behaviour of Drp1 puncta during division (Extended Data Fig. 4c–e). Conversely, Dyn2 puncta segregate to only one of the daughter mitochondria after division (Fig. 2f, g). The differences in puncta dynamics between Drp1 and Dyn2 display a strong resemblance to the division machineries that collaborate to drive organelle division in the alga *Cyanidioschyzon merolae*. *C. merolae* relies on sequential constriction events, first mediated by the FtsZ-ring machinery, then by an organelle-specific dynamin, the Drp1-related Dnm1 and the Dyn2-related Dnm2, to divide mitochondria and chloroplasts, respectively^{21,22}. The behaviour of FtsZ-ring puncta mirrors mammalian Drp1 puncta dynamics during mitochondrial division, whereas the puncta of *C. merolae* dynamins act in a similar way to mammalian Dyn2 puncta^{21,22}. Taken together, these data support a direct role for Dyn2 in mitochondrial division, downstream of Drp1 constriction.

Dyn2 puncta appear at Drp1-marked mitochondrial constriction sites just before division, which suggests that Dyn2 functions downstream of Drp1. To test this further, we evaluated the occupancy of Drp1 puncta along mitochondria (puncta μm^{-2}). The percentage of mitochondria-associated Drp1 puncta was not significantly different between control and Dyn2-depleted cells (Fig. 3a). The endoplasmic-reticulum-labelled (via GFP-Sec61β) mitochondrial constrictions have been identified as a pre-Drp1 step because these constrictions formed in the absence of Drp1⁸. We could also qualitatively observe that Drp1 puncta still localize to endoplasmic-reticulum-labelled mitochondrial constrictions in Dyn2-depleted cells (Fig. 3b, c). Taken together, these data further support a role for Dyn2 in mitochondrial division that is downstream of Drp1 recruitment.

We envisioned that the division process in Dyn2-depleted cells would fail at a step after Drp1 recruitment, possibly leading to the accumulation of Drp1-marked stalled ‘super-constrictions’ that are unable to divide. To investigate this, we activated mitochondrial division in control and Dyn2-depleted cells by acute addition of the Ca^{2+} chelator, BAPTA, which has been previously used to induce Drp1-dependent division⁸. In control cells, treatment with BAPTA induced Drp1-marked mitochondrial division, with Drp1 puncta splitting to maintain association with each daughter mitochondrion (Fig. 3d and Supplementary Video 4). By contrast, BAPTA-treated Dyn2-depleted cells displayed Drp1-labeled constrictions that appeared destined for mitochondrial division but that failed to divide. Line-scan analysis of these dynamic events identified a novel post-Drp1 intermediate ‘super-constriction’, characterized as the transient splitting of Drp1 puncta coupled with a coincidental increase in the length of a constriction along the longitudinal axis (Fig. 3e and Supplementary Video 5). These data reveal the existence of a transient post-Drp1 super-constriction step that occurs between Drp1 constriction and membrane fission and requires Dyn2 to proceed.

Compelling *in vitro* experiments highlight the mechanical differences between Drp1 and classical dynamins^{7,10,11}. Drp1 constricted liposomes to 31 ± 5 nm, whereas Dyn1 was shown to assemble on lipid tubules with a diameter of 52 ± 4 nm, inducing fission of the tubules upon GTP hydrolysis^{7,10}. Additionally, super-resolution live-cell imaging studies have identified mitochondrial division intermediates as constrictions with extended length measurements (102 ± 16 nm); however,

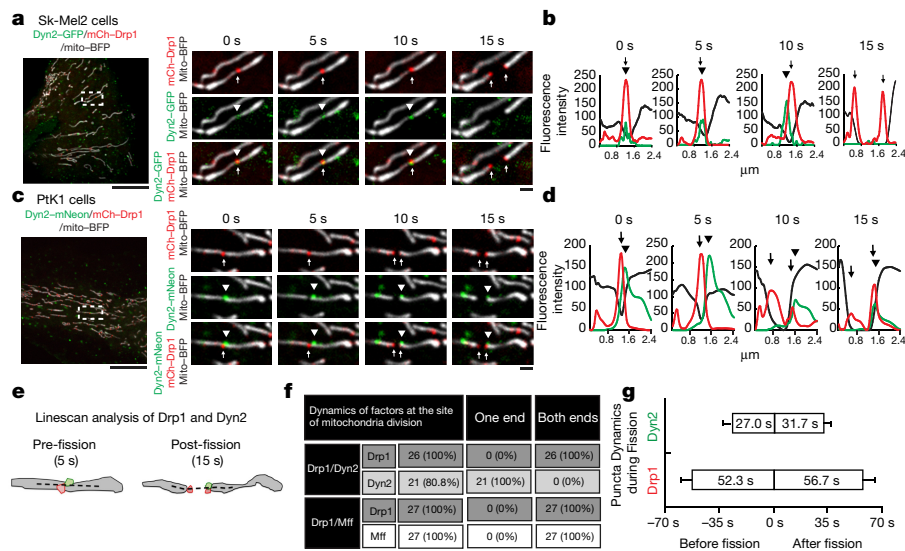


Figure 2 | Dynamics of Dyn2 recruitment during mitochondrial division. **a–d** Mitochondrial division was imaged live in Sk-Mel2 cells (14 events from 144 cells; **a, b**) and PtK cells (26 events from 254 cells; **c, d**). **a, c**, Highlighted insets show the spatiotemporal dynamics of Drp1 (arrows) and Dyn2 (arrowheads) before and during mitochondrial division. **b, d**, Line-scan analysis of mean fluorescence intensity verifying Drp1 and Dyn2 dynamics. **e, f**, A cartoon schematic depicting the frame before division that determines the presence of division factors (Dyn2 and Drp1) at the site of division, and the frame after division that determines the interactions of division factors with the ends of daughter mitochondria (**e**). These are summarized in **f**. The dashed line in **e** represents area analysed by line scan. **g**, Lifetimes of Drp1 and Dyn2 puncta on mitochondria before and after division were analysed from 21 division events. Error bars represent s.e.m. **b, d**, Scale bars for whole-cell (**a, c**; left) panels, 10 μm ; for inset panels (**a, c**; right), 1 μm .

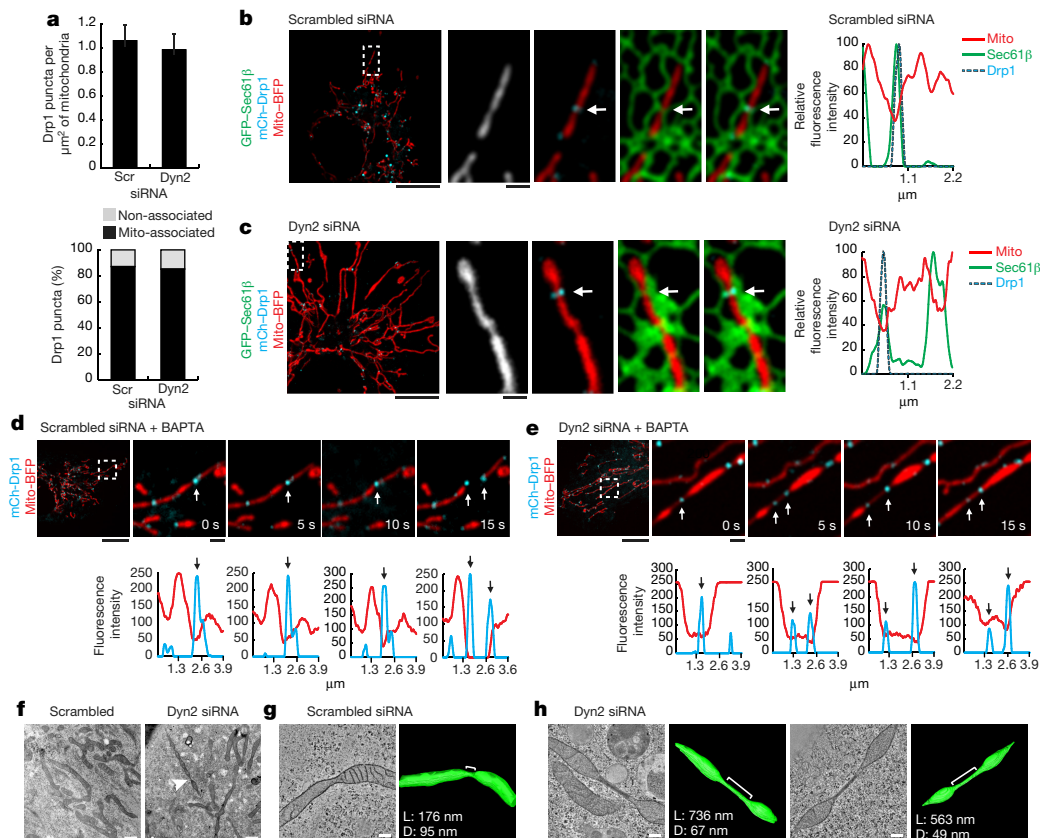


Figure 3 | Dyn2-depleted cells reveal dynamic mitochondrial super-constrictions. **a**, Analysis of mitochondrial association of Drp1 puncta; $n = 12$ cells for both scrambled siRNA and Dyn2 siRNA treatments. $*P < 0.01$ statistical significance calculated by paired t -test. Error bars represent s.e.m. **b, c** Examples of Drp1-marked constrictions at endoplasmic-reticulum-mitochondria contacts (arrows) in scrambled (**b**; $n = 12$) or Dyn2 siRNA-treated (**c**; $n = 12$) cells. The mean fluorescence intensity of each factor was derived from line scans and plotted (graphs, right). **d, e**, Representative images of scrambled (**d**; Supplementary Video 4, 11 events from 31 cells) versus Dyn2 (**e**; Supplementary Video 5, 10 events from 34 cells) siRNA-treated cells were captured following acute BAPTA treatment. The mean fluorescence intensity of Drp1 and mitochondria were derived from line scans and plotted for each time point

(graphs, below). **f**, Representative electron micrographs of scrambled ($n = 40$ cells) and Dyn2 siRNA-treated ($n = 53$ cells) cells captured from serial thin sections revealed severe mitochondrial constrictions (arrowhead) in Dyn2 siRNA-treated cells. **g–h**, Representative electron tomographs and corresponding 3D models of mitochondrial constrictions (<100 nm diameter) observed in scrambled (4 constrictions from 12 sections; **g**) or (**h**) Dyn2 siRNA-treated (10 constrictions from 12 sections) cells. In control cells, the mean diameter (D) and length (L) are 77 ± 20 nm and 143 ± 61 nm, respectively. Mitochondrial constrictions in Dyn2-depleted cells displayed a diameter of 55 ± 12 nm and length of 358 nm with a minimum and maximum of 145 and 771 nm, respectively. Scale bars, $10 \mu\text{m}$ (**b, c, e**; leftmost images), $1 \mu\text{m}$ (**b, c, e**; right four images), $1,000$ nm (**f**), 200 nm (**g, h**).

Drp1 was not captured in these studies²³. Our live-cell studies have seemingly captured and stalled Drp1-marked division intermediates through the depletion of Dyn2 under division-promoting conditions—an event characterized as the transient splitting of Drp1 puncta that coincides with an increased constriction length (Fig. 3d, e).

We next used electron microscopy and 3D tomography to obtain ultrastructural information about mitochondrial constrictions in Dyn2-depleted cells. Dual-tilt tomographs of control and Dyn2-siRNA-treated COS-7 cells were captured from 300-nm-thick sections and reconstructed to allow for the precise measurement of the length and diameter of mitochondrial constrictions (Fig. 3f–h). We aimed to characterize mitochondrial constrictions with diameters that are suitable for Drp1 assembly (that is, <100 nm) in scrambled- and Dyn2-siRNA-treated cells and could resolve four and ten mitochondrial constrictions that fit this criteria, respectively. The constrictions in scrambled-siRNA-treated cells had a mean diameter of 77 ± 20 nm and a mean length of 143 ± 61 nm (Fig. 3g). By contrast, constrictions in Dyn2-siRNA-treated cells were narrower (mean diameter of 55 ± 12 nm) and longer (mean length of 358 nm, with a minimum and maximum of 145 and 771 nm, respectively; Fig. 3h). Together, these data display stalled and elongated super-constrictions in Dyn2-depleted cells, with diameters that could be generated by Drp1 GTP hydrolysis and are suitable for Dyn2 assembly^{7,10,11}. Therefore, we propose a model of sequential constriction that includes multiple dynamins, whereby Drp1 constricts mitochondria to a diameter that is conducive for Dyn2 assembly, which then drives further membrane constriction and complete fission.

Mitochondria have a critical role in apoptotic cell death¹². Cells treated with staurosporine (STS), a global kinase inhibitor that induces apoptosis, exhibit Drp1-dependent mitochondrial fragmentation, cytochrome *c* release from, and Bax recruitment to, mitochondria^{14,24}. Following STS treatment, we observed an increase in Drp1 localization to Mff-marked mitochondrial constrictions, as well as increased localization of Dyn2 to Drp1-marked constrictions (Extended Data Fig. 5). Furthermore, control cells treated with STS exhibited fragmented mitochondria, whereas Drp1- and Dyn2-depleted cells maintained an elongated mitochondrial network after 1.5 h of STS treatment (Fig. 4a, b). Immunofluorescence and fractionation studies of cells treated with STS for 1.5 and 5 h showed that Drp1 or Dyn2 depletion delays the release of cytochrome *c* from mitochondria (Fig. 4c and Extended Data Fig. 6) and slows the recruitment of Bax to mitochondria (Fig. 4c), despite accelerating Bax activation (Extended Data Fig. 7). Thus, Dyn2 also plays a critical role in apoptosis-induced, Drp1-dependent mitochondrial fragmentation, supporting the hypothesis that Drp1 and Dyn2 are involved in the same mitochondrial pathway.

Our results show that mitochondrial division is a Dyn2-dependent process, and highlight a conserved overlap of dynamins with roles at both mitochondria and plasma membrane. Previous studies in *C. merolae* and other organisms have suggested that mitochondrial dynamins in other organisms have a dual role, acting in both mitochondrial division and endocytosis^{6,21,25}. Deletion of the only dynamin gene in *Trypanosoma brucei* caused defects in mitochondrial division, endocytosis and cytokinesis²⁵. This versatility has been preserved in mammalian Dyn2, which has already been shown to function in endocytosis and cytokinesis^{13,17}. Dyn2 is recruited to endocytic sites by SH3-domain proteins and, once there, its affinity for both membrane curvature^{11,18} and anionic phospholipids²⁶, specifically phosphoinositide (4,5)-biphosphate, accelerates its assembly into a functional fission machine. Adaptor proteins and lipid environments also play significant roles in mitochondrial dynamics^{27–29}. Future studies will be aimed at identifying the recruitment factors and mitochondrial lipid interactions of Dyn2.

We propose that mammalian mitochondrial division has integrated several machineries that proceed in multiple constriction steps. First, endoplasmic-reticulum- and actin-driven constrictions permit Drp1 assembly; second, Drp1-mediated constriction facilitates Dyn2

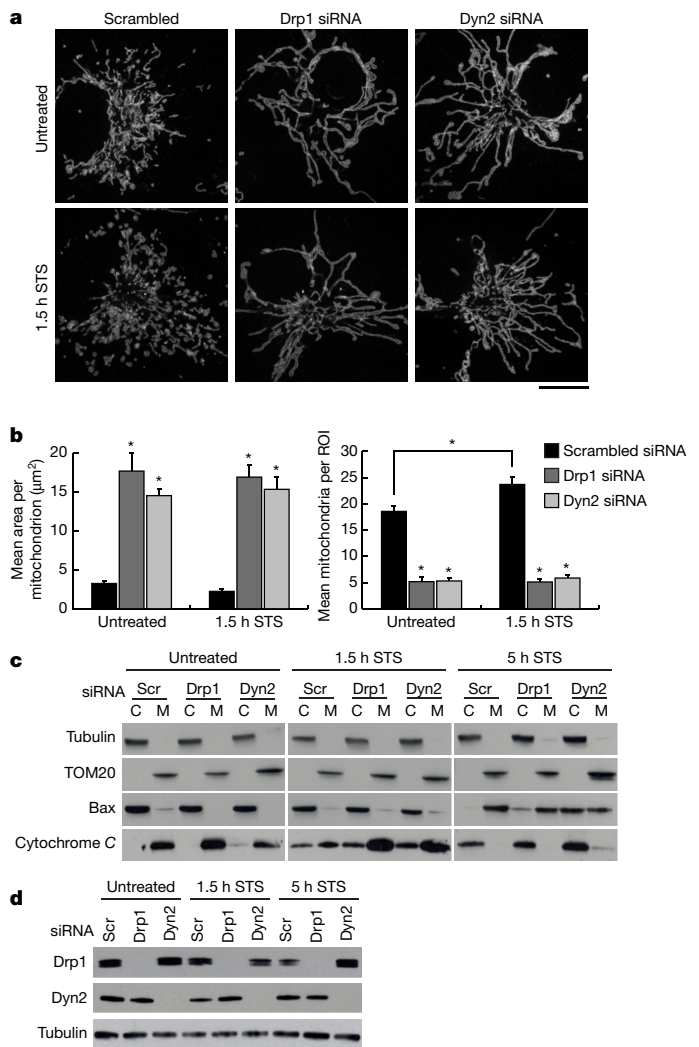


Figure 4 | Dyn2 is required for STS-induced mitochondrial fragmentation. **a**, Representative images of TOM20 immunofluorescence in $n = 41$, 42 scrambled siRNA cells; $n = 45$, 43 Drp1 siRNA; and $n = 45$, 43 Dyn2 siRNA without and with STS treatment, respectively. Scale bars, 10 μm. **b**, Mitochondrial morphology was quantified for mean mitochondrial area per mitochondrion, and mean mitochondria per ROI. Error bars represent s.e.m. * $P < 0.01$ by ANOVA, obtained from three biological replicates. **c**, Cytochrome *c* release and Bax recruitment were evaluated by immunoblot of cytosol (C) and membrane (M) fractions. α-Tubulin verified cytosolic fractions and TOM20 verified membrane fractions. **d**, Drp1 and Dyn2 depletion were confirmed by immunoblot of post-nuclear supernatants.

assembly; and third, Dyn2-mediated constriction induces membrane fission to complete division.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 November 2015; accepted 24 October 2016.

Published online 31 October 2016.

- Goldstein, S., Moerman, E. J. & Porter, K. High-voltage electron microscopy of human diploid fibroblasts during ageing *in vitro*: morphometric analysis of mitochondria. *Exp. Cell Res.* **154**, 101–111 (1984).
- Legesse-Miller, A., Massol, R. H. & Kirchhausen, T. Constriction and Dnm1p recruitment are distinct processes in mitochondrial fission. *Mol. Biol. Cell* **14**, 1953–1963 (2003).
- Kim, H.-W. *et al.* Efficient and accurate analysis of mitochondrial morphology in a whole cell with a high-voltage electron microscopy. *J. Electron Microsc. (Tokyo)* **61**, 127–131 (2012).

4. Otsuga, D. *et al.* The dynamin-related GTPase, Dnm1p, controls mitochondrial morphology in yeast. *J. Cell Biol.* **143**, 333–349 (1998).
5. Smirnova, E., Griparic, L., Shurland, D. L. & van der Bliek, A. M. Dynamin-related protein Drp1 is required for mitochondrial division in mammalian cells. *Mol. Biol. Cell* **12**, 2245–2256 (2001).
6. Purkanti, R. & Thattai, M. Ancient dynamin segments capture early stages of host-mitochondrial integration. *Proc. Natl Acad. Sci. USA* **112**, 2800–2805 (2015).
7. Yoon, Y., Pitts, K. R. & McNiven, M. A. Mammalian dynamin-like protein DLP1 tubulates membranes. *Mol. Biol. Cell* **12**, 2894–2905 (2001).
8. Friedman, J. R. *et al.* ER tubules mark sites of mitochondrial division. *Science* **334**, 358–362 (2011).
9. Prudent, J. & McBride, H. M. Mitochondrial dynamics: ER actin tightens the Drp1 noose. *Curr. Biol.* **26**, R207–R209 (2016).
10. Sweitzer, S. M. & Hinshaw, J. E. Dynamin undergoes a GTP-dependent conformational change causing vesiculation. *Cell* **93**, 1021–1029 (1998).
11. Roux, A. *et al.* Membrane curvature controls dynamin polymerization. *Proc. Natl Acad. Sci. USA* **107**, 4141–4146 (2010).
12. Wang, C. & Youle, R. J. The role of mitochondria in apoptosis. *Annu. Rev. Genet.* **43**, 95–118 (2009).
13. Ferguson, S. M. & De Camilli, P. Dynamin, a membrane-remodelling GTPase. *Nat. Rev. Mol. Cell Biol.* **13**, 75–88 (2012).
14. Frank, S. *et al.* The role of dynamin-related protein 1, a mediator of mitochondrial fission, in apoptosis. *Dev. Cell* **1**, 515–525 (2001).
15. Gomes, L. C., Di Benedetto, G. & Scorrano, L. During autophagy mitochondria elongate, are spared from degradation and sustain cell viability. *Nat. Cell Biol.* **13**, 589–598 (2011).
16. Taguchi, N., Ishihara, N., Jofuku, A., Oka, T. & Mihara, K. Mitotic phosphorylation of dynamin-related GTPase Drp1 participates in mitochondrial fission. *J. Biol. Chem.* **282**, 11521–11529 (2007).
17. Thompson, H. M., Skop, A. R., Euteneuer, U., Meyer, B. J. & McNiven, M. A. The large GTPase dynamin associates with the spindle midzone and is required for cytokinesis. *Curr. Biol.* **12**, 2111–2117 (2002).
18. Morlot, S. & Roux, A. Mechanics of dynamin-mediated membrane fission. *Annu. Rev. Biophys.* **42**, 629–649 (2013).
19. Bui, H. T. & Shaw, J. M. Dynamin assembly strategies and adaptor proteins in mitochondrial fission. *Curr. Biol.* **23**, R891–R899 (2013).
20. Doyon, J. B. *et al.* Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nat. Cell Biol.* **13**, 331–337 (2011).
21. Nishida, K. *et al.* Dynamic recruitment of dynamin for final mitochondrial severance in a primitive red alga. *Proc. Natl Acad. Sci. USA* **100**, 2146–2151 (2003).
22. Miyagishima, S. Y. *et al.* A plant-specific dynamin-related protein forms a ring at the chloroplast division site. *Plant Cell* **15**, 655–665 (2003).
23. Shim, S.-H. *et al.* Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes. *Proc. Natl Acad. Sci. USA* **109**, 13978–13983 (2012).
24. Otera, H. *et al.* Mff is an essential factor for mitochondrial recruitment of Drp1 during mitochondrial fission in mammalian cells. *J. Cell Biol.* **191**, 1141–1158 (2010).
25. Chanez, A.-L., Hehl, A. B., Engstler, M. & Schneider, A. Ablation of the single dynamin of *T. brucei* blocks mitochondrial fission and endocytosis and leads to a precise cytokinesis arrest. *J. Cell Sci.* **119**, 2968–2974 (2006).
26. Burger, K. N. J., Demel, R. A., Schmid, S. L. & de Kruijff, B. Dynamin is membrane-active: lipid insertion is induced by phosphoinositides and phosphatidic acid. *Biochemistry* **39**, 12485–12493 (2000).
27. Otera, H., Miyata, N., Kuge, O. & Mihara, K. Drp1-dependent mitochondrial fission via MiD49/51 is essential for apoptotic cristae remodeling. *J. Cell Biol.* **212**, 531–544 (2016).
28. Osellame, L. D. *et al.* Cooperative and independent roles of Drp1 adaptors Mff and MiD49/51 in mitochondrial fission. *J. Cell Sci.* jcs.185165 (2016). <http://dx.doi.org/10.1242/jcs.185165>
29. Osman, C., Voelker, D. R. & Langer, T. Making heads or tails of phospholipids in mitochondria. *J. Cell Biol.* **192**, 7–16 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Ozzello for sectioning cells for electron microscopy analysis; A. Cheng, J. Ryoo and D. G. Drubin for providing the genome-edited Dyn2–GFP Sk-Mel2 cell line; J. R. Friedman, D. Dambournet, and K. C. Cook for critical reading of the manuscript; and A. Hoenger and M. West for helpful discussions. Electron microscopy was performed at the University of Colorado, Boulder electron microscopy services core facility. Super-resolution microscopy in the Department of Molecular, Cellular, and Developmental Biology was made possible by equipment supplements R01 GM79097 (D. Xue) and P01 GM105537 (M. Winey). This work is supported by grants from the National Institutes of Health to G.K.V. (GM083977), J.E.L. (F32CA174158), L.M.W. (F32GM116371) and to H.W. (T32GM08759).

Author Contributions J.E.L. and G.K.V. contributed to experimental design. L.M.W. and C.P. performed the experiments in Fig. 3f–h; H.W. performed the experiments in Extended Data Fig. 4d–f; J.E.L. conducted the remainder of the experiments, data analysis, and figure composition. J.E.L. and G.K.V. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.K.V. (gia.voeltz@colorado.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

DNA plasmids and small interfering RNA oligonucleotides. GFP–Sec61 β (ref. 8), mCherry–Drp1 isoform 3 (NM_005690)⁸, GFP–Mff (NM_020194)⁸ and Mito–BFP⁸ were previously described. Dynamin-2 isoform 3 (NM_004945) was cloned from HeLa cDNA and inserted into the XhoI/EcoRI sites of pAcGFP–N1 (Clontech) and pmNeonGreen–N1 to give Dyn2–GFP and Dyn2–mNeon, respectively. Site-directed mutagenesis was used to generate six silent mutations per siRNA seed region in Dyn2–mNeon, creating siRNA-resistant Dyn2–mNeon mRNA that translates wild-type Dyn2–mNeon. Subsequent site-directed mutagenesis reactions were performed to generate siRNA-resistant, dominant-negative Dyn2(K44A)–mNeon. The siRNA-resistant Dyn2(Δ PH)–mNeon mutant was constructed by first, linearizing the siRNA-resistant Dyn2–mNeon by PCR with primers that exclude the PH-domain sequence, followed by Gibson assembly to circularize the construct. The siRNA-resistant Dyn2(Δ PRD)–mNeon mutant was constructed by PCR amplification of siRNA-resistant Dyn2–mNeon with primers that exclude the C-terminal proline-rich domain, then digested with XhoI/KpnI and ligated into the mNeon–N1 vector.

Oligonucleotides for Drp1 siRNA were synthesized by Qiagen (SI02661365) against the target sequence: 5′-CAGGAGCCAGCTAGATATTA-3′. A pool of two siRNAs was used to substantially deplete Dyn2. Two oligonucleotides for Dyn2 siRNA were synthesized by GE Healthcare Dharmacon against the two target sequences, Dyn2 siRNA#1 (J-004007-06-0005): 5′-GAGCGAATCGTCACCACTT-3′ and Dyn2 siRNA#2 (J-004007-08-0005): 5′-GAGCGAATCGTCACCACTT-3′. Silencer Negative Control #1 siRNA (AM4635) was synthesized by Ambion. Oligonucleotides for AP-2 siRNA against the target sequence: 5′-GAGCATGTGCACGCTGGCCAGCT-3′ were a gift from S. M. DiPietro.

Cell culture, transfection, and drug treatment. Genome-edited Dyn2–GFP skin melanoma cells (Sk-Mel2/Dyn2–GFP) were a gift from D. G. Drubin. COS-7 cells (ATCC-CRL-1651), HeLa cells (ATCC-CCL-2), PtK1 cells (ATCC-CRL-6493) were verified by cytochrome *c* oxidase subunit I (CO1) interspecies assay and CCL-2 HeLa cells were further verified by short tandem repeat testing. Cell lines were tested for Mycoplasma contamination by ATCC at the time of purchase. Cells were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. COS-7, HeLa, and PtK1 cells were seeded on plastic-bottomed dishes at 1×10^5 cells ml⁻¹ about 16 h before transfection, whereas Sk-Mel2 Dyn2–GFP cells were plated directly onto 35-mm glass-bottomed microscope dishes (Mattek) at 0.5×10^5 cells ml⁻¹ about 40 h before transfection. Plasmid transfections were performed with 2.5 μ l Lipofectamine 2000 or 3000 (Invitrogen) per ml OPTI-MEM media (Invitrogen) for around 5 h, followed by splitting of COS-7, HeLa, and PtK1 cells to 35-mm glass-bottom microscope dishes (MatTek) at a density of 1×10^5 cells ml⁻¹. Roughly 16–24 h after transfection, cells were imaged in pre-warmed 37°C Fluorobrite DMEM (Invitrogen) supplemented with 10% FBS. For all experiments, the following amounts of DNA were transfected per ml: 150 ng GFP–Sec61 β ; 40 ng mito–BFP (amino acids 1–22 of cytochrome *c* oxidase subunit IV from *Saccharomyces cerevisiae* fused to BFP); 40 ng mCherry–Drp1; 50 ng Dyn2–GFP variants; 25 ng Dyn2–mNeon; 40 ng GFP–Mff. RNAi transfections were performed as above, with an additional transfection step: cells were seeded as above and transfected with 2.5 μ l Dharmafect (Dharmacon) and 40 nM siRNA oligonucleotides, split 1:2 ~24 h post-transfection, and re-transfected ~24 h later with 5 μ l Lipofectamine 2000, 40 nM siRNA oligonucleotides, and appropriate plasmid DNA. For knockdown and rescue studies, COS-7 cells were transfected with Dyn2 siRNA to deplete endogenous Dyn2 mRNA. A second round of transfection was performed 2 days later with Dyn2 siRNA and DNA plasmids encoding mNeon alone or siRNA-resistant variants of Dyn2–mNeon. Simultaneous depletion of endogenous Dyn2 and unabated expression of exogenous siRNA-resistant Dyn2–mNeon variants was verified by immunoblot analysis.

In BAPTA experiments, cells were treated with 10 μ M BAPTA-AM (Calbiochem) in Fluorobrite DMEM during imaging at 37°C. In STS treatment experiments, cells were first treated with 75 μ M of the caspase inhibitor z-VAD-fmk (BD Biosciences) for 4 h followed by 1 μ M staurosporine (Sigma-Aldrich) for 1.5 or 5 h, followed by fractionation or fixation with 4% paraformaldehyde, 0.5% glutaraldehyde in PBS in preparation for immunofluorescence. For fractionation experiments, cells were collected from 10-cm plates, washed with ice-cold PBS, and resuspended in 1 ml of isolation buffer (225 mM mannitol, 75 mM sucrose, 30 mM Tris-HCl pH 7.4, 0.1 mM EGTA). The plasma membrane was ruptured by Dounce homogenization. Cell lysates were spun at 600 g for 5 min at 4°C to pellet and remove nuclei and debris. The post-nuclear supernatant (PNS) was

collected and spun at 100,000 g to pellet membranes. The supernatant was collected and represented the cytosolic fraction. The pellet (membrane fraction) was resuspended in solubilization buffer (50 mM HEPES, 2.5 mM MgCl₂, 200 mM KCl, 5% glycerol, 1% Triton X-100). Cytosolic and membrane fractions were then treated with trichloroacetic acid (TCA) to precipitate proteins, which were then resuspended in equal volumes of 2 \times Laemmli buffer.

Protein levels in whole-cell lysates and cell fractions were assayed by western blot with polyclonal rabbit antibodies against Dyn2 (ab3457, Abcam), phospho-Ser637–Drp1 (4867S, Cell Signaling Tech), phospho-Ser616–Drp1 (3455S, Cell Signaling Tech), Mff (17090-1-AP, Protein Tech), Mfn2 clone 4H8 (WH0009927M3, Sigma-Aldrich), α -tubulin (ab18251, Abcam), TOM20 (sc-11415, Santa Cruz Biotech), Bax (sc-493, Santa Cruz Biotech), GAPDH (G9545, Sigma-Aldrich), and the monoclonal mouse antibody against AP-1/2 (sc-17771, Santa Cruz Biotech.), Drp1 (ab56788, Abcam), dynamin (610245, BD Biosciences), cytochrome *c* (sc-13560, Santa Cruz Biotech) and Opa1 (612606, BD Biosciences). Immunofluorescence was accomplished using monoclonal antibodies against cytochrome *c* clone 6H2. B4 (12963S, Cell Signaling Tech) and Bax6A7 (B8429, Sigma Aldrich), and the polyclonal rabbit antibody against TOM20 (sc-11415, Santa Cruz Biotech).

Microscopy. Imaging was performed with an inverted fluorescence microscope (TE2000-U; Nikon) equipped with an electron-multiplying charge-coupled device camera (Cascade II; Photometrics) and a Yokogawa spinning-disc confocal system (CSU-Xm2; Nikon). Live-cell imaging was performed at 37°C. Images were captured with a 100 \times NA 1.4 oil objective and acquired using MetaMorph (version 7.0; MDS Analytical Technologies).

Super-resolution microscopy (SIM) was performed with an inverted fluorescence microscope (TI-E; Nikon) equipped with an electron-multiplying charge-coupled device camera (iXon3 897; Andor). z-stack images were captured every 200 nm and reconstructed using NIS elements with SIM (Nikon).

Analyses of mitochondrial morphology. COS-7 and HeLa cells were treated with siRNA to deplete targeted proteins, and then mitochondria were imaged either live with mito–BFP or fixed with immunofluorescence of polyclonal rabbit TOM20 antibody (sc-11415, Santa Cruz Biotechnologies) probed with Alexa fluor 488 donkey anti-rabbit antibody (A21206, Invitrogen). To ensure that mitochondria present through multiple planes of the cell could be assayed, z-stacks with a thickness of 2.4 μ m were acquired of each cell, and consisted of 12 serial images that were each spaced by 0.2 μ m. The morphology of mitochondria within a resolvable 230 μ m² ROI, which was peripheral to the microtubule-organizing centre, was assessed using various tools in ImageJ (National Institutes of Health). First, maximum intensity projections were generated from z-stacks followed by automated 8-bit Otsu Thresholding. Lastly, thresholded images were evaluated using the 'Analyze Particles' function in ImageJ to obtain the number of mitochondrial fragments and the area of each fragment per ROI.

Capturing mitochondrial division and analysing mitochondrial constrictions. Dynamic events and interactions, such as mitochondrial division and constriction, were imaged live with fluorescently tagged proteins. Time-lapse videos were acquired over the course of 2 min with each channel captured every 5 s. Exposure times ranged between 200–300 ms in each channel. Dyn2–GFP, Dyn2–mNeon, GFP–Mff, mCherry–Drp1, and mito–BFP were imaged in mitochondrial division studies. Drp1-dependent fission events were resolvable when the splitting of a mitochondrion into two distinct mitochondria coincided with the splitting of a Drp1 punctum, which results in Drp1 puncta labelling the ends of each fission product. Division events were further verified by line-scan analysis using ImageJ. A segmented line was drawn along the length of the mitochondria, and then the fluorescence intensity of all channels was measured along the length of the line.

COS-7 cells expressing GFP–Sec61 β , mCherry–Drp1, and mito–BFP were used to investigate mitochondrial constrictions. The diffuse cytosolic staining of Drp1 was threshold-adjusted from Drp1 puncta using the rolling-ball background subtraction method (radius, 8 pixels or 0.304 μ m)³⁰. Mitochondrial-associated Drp1 puncta were identified using the 'Colocalization Highlighter' ImageJ plugin, then counted using the 'Analyze Particles' function in ImageJ. The occupancy of Drp1 puncta along mitochondria (puncta μ m⁻²) and the percentage of mitochondria-associated Drp1 puncta were quantitated. Qualitatively, endoplasmic-reticulum-marked constrictions were analysed with line-scans that were performed in ImageJ. Endoplasmic-reticulum-marked constrictions were identified by acute decreases in mito–BFP fluorescence that coincided with GFP–Sec61 β and/or mCherry–Drp1 fluorescence peaks. An additional criteria used to identify constrictions in time-lapse video was to observe whether the intersection of these two organellar markers was dynamically linked in movement.

Mitochondrial-associated Drp1 puncta were identified using the 'Colocalization Highlighter' ImageJ plugin, then counted using the 'Analyze Particles' function in ImageJ. The occupancy of Drp1 puncta along mitochondria (puncta μ m⁻²) and the percentage of mitochondria-associated Drp1 puncta were quantitated.

Qualitatively, endoplasmic-reticulum-marked constrictions were analysed with line scans that were performed in ImageJ. Endoplasmic-reticulum-marked constrictions were identified by acute decreases in mito-BFP fluorescence that coincided with GFP-Sec61 β and/or mCherry-Drp1 fluorescence peaks. An additional criteria used to identify constrictions in time-lapse video was to observe whether the intersection of these two organellar markers was dynamically linked in movement.

Electron microscopy. Sapphire discs (3 mm \times 0.05 mm, Swiss Micro Technology) were used as a platform to grow cells and were coated with carbon and overlaid with a gold finder pattern using EMS950X Turbo Evaporator (Electron Microscopy Sciences) as described previously³¹. Discs were coated with poly-L-lysine (Sigma, 0.01% solution), air-dried then glow-discharged (EMITECH Ltd model K 100 \times Glow Discharge) and immobilized to the bottom of 35-mm tissue culture dishes with 10 μ l of poly-L-lysine that was allowed to dry before ultraviolet radiation.

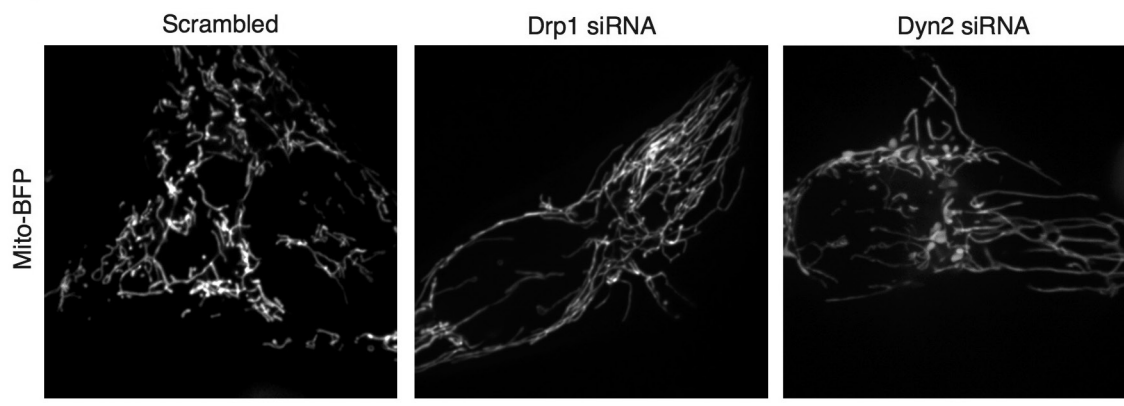
Cells were transfected with siRNA as described above and plated (1×10^5 cells ml⁻¹) on sapphire discs 24 h before freezing. Discs were removed from culture, swirled in a cryo-protectant medium (DMEM, 10% FBS, 150 mM mannitol, and 2% sucrose) and transferred to aluminium freezing planchettes (Wohlgend, Switzerland) that were coated with hexadecane (Sigma), then rapidly frozen using the Wohlgend Compact 02 high-pressure freezer. The frozen discs were kept in liquid nitrogen, then transferred into cryo-tubes (Nalgene-Nunc) for freeze-substitution in 1% osmium tetroxide, 0.1% uranyl acetate and 2% water in glass-distilled acetone (EMS). The tubes were then transferred to an aluminium block pre-chilled to -80°C and gradually warmed to -30°C overnight (~ 16 h) in a -30°C freezer and then warmed to room temperature (around 4 h). The samples were washed 3 times with acetone (glass-distilled; EMS). Following the 3rd wash, discs were oriented cell side up and covered in 25% epon-araldite resin in acetone. Cells were gently infiltrated with increasing concentrations of epon-araldite resin (25% for 4 h, 50% for 12–16 h, 75% for 16 h, and 100% for 24 h). Finally, discs were changed into a solution of 100% epon-araldite plus accelerator (DMP-30) before being placed in mounting chambers and polymerized at 60°C for 48 h (refs 31,32).

Embedded cells were re-mounted onto resin stubs and sectioned at 70- or 300-nm thickness (thin section or tomography, respectively), then collected onto a formvar-coated slot grid and post-stained using 2% uranyl acetate in water (8 min) and Reynolds lead citrate (3 min) as described previously^{31–33}. Fiducial markers (BB International) for tomography were added by placing both sides of grids on a drop of 15 nm gold colloid solution. Cellular regions were selected and imaged with a Tecnai T12 Spirit 120 kV for 70 nm sections. For tomography, 300-nm sections were imaged on the Tecnai TF30 300 kV. Dual axis tilts $\pm 60^\circ$ in increments of 1.5° were recorded around the orthogonal axes. Serial electron microscopy was used for image acquisition and each tilt series was reconstructed using the IMOD software^{34–36}. Tomographic reconstructions were modelled by manual contour tracing using IMOD and constriction length and width measurements were taken at the widest view of the constriction using slicer (IMOD) which allows for manipulation of the reconstruction in the x , y and z dimensions.

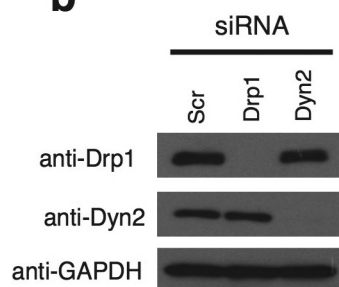
Data availability. The source data for all gels and blots are provided as a Supplementary Figure in the online version of the paper. All other data that support the findings of this study are available from the authors upon reasonable request.

30. Ji, W. K., Hatch, A. L., Merrill, R. A., Strack, S. & Higgs, H. N. Actin filaments target the oligomeric maturation of the dynamin GTPase Drp1 to mitochondrial fission sites. *eLife* **4**, e11553 (2015).
31. McDonald, K. *et al.* "Tips and tricks" for high-pressure freezing of model systems. *Methods Cell Biol.* **96**, 671–693 (2010).
32. Schiel, J. A. *et al.* Endocytic membrane fusion and buckling-induced microtubule severing mediate cell abscission. *J. Cell Sci.* **124**, 1411–1424 (2011).
33. O'Toole, E. T., Giddings, T. H., Jr & Dutcher, S. K. Understanding microtubule organizing centers by comparing mutant and wild-type structures with electron tomography. *Methods Cell Biol.* **79**, 125–143 (2007).
34. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
35. Mastronarde, D. N. Dual-axis tomography: an approach with alignment methods that preserve resolution. *J. Struct. Biol.* **120**, 343–352 (1997).
36. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

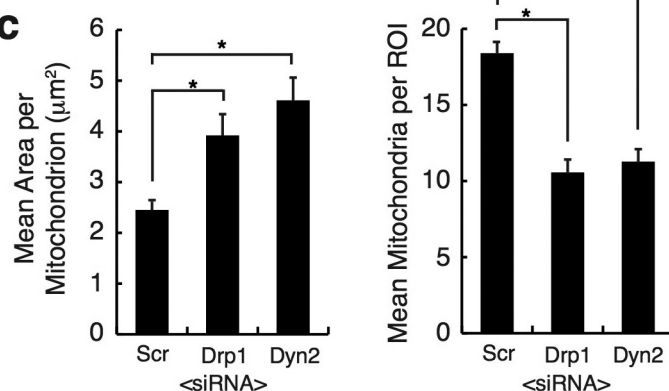
a HeLa cells



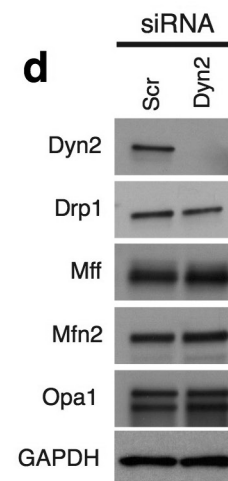
b



c

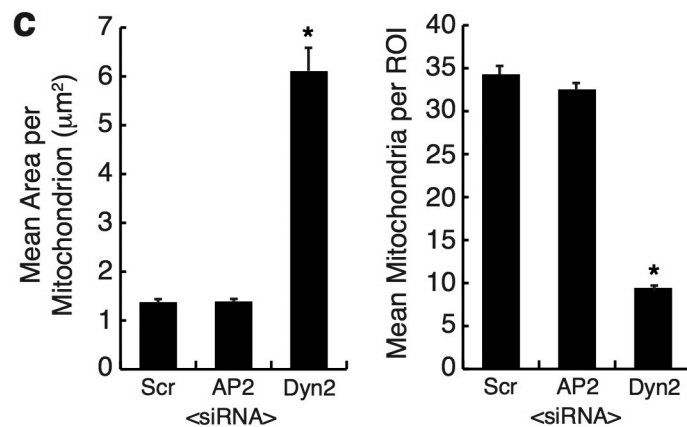
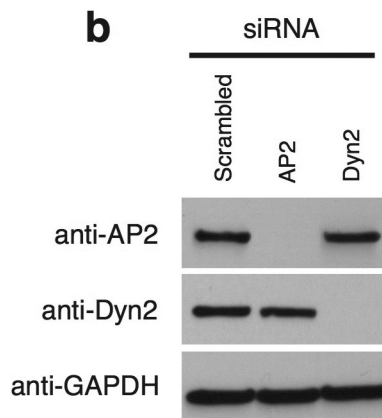
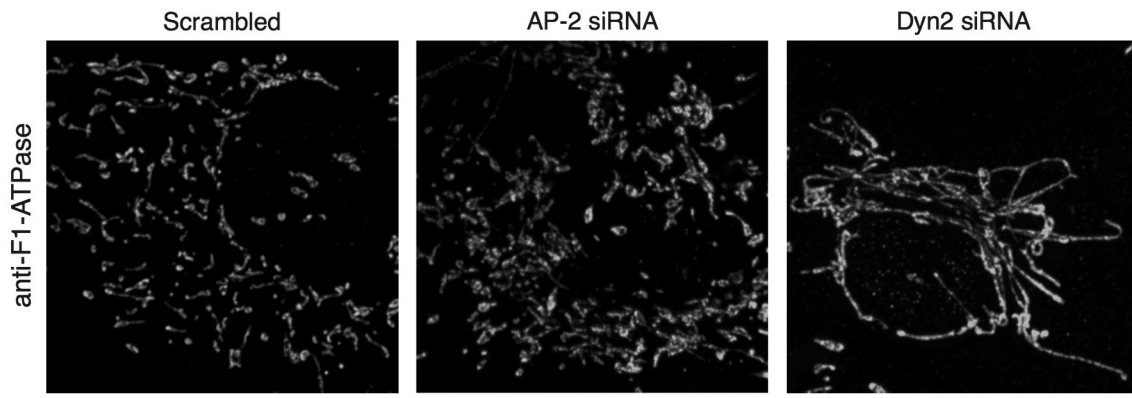


d



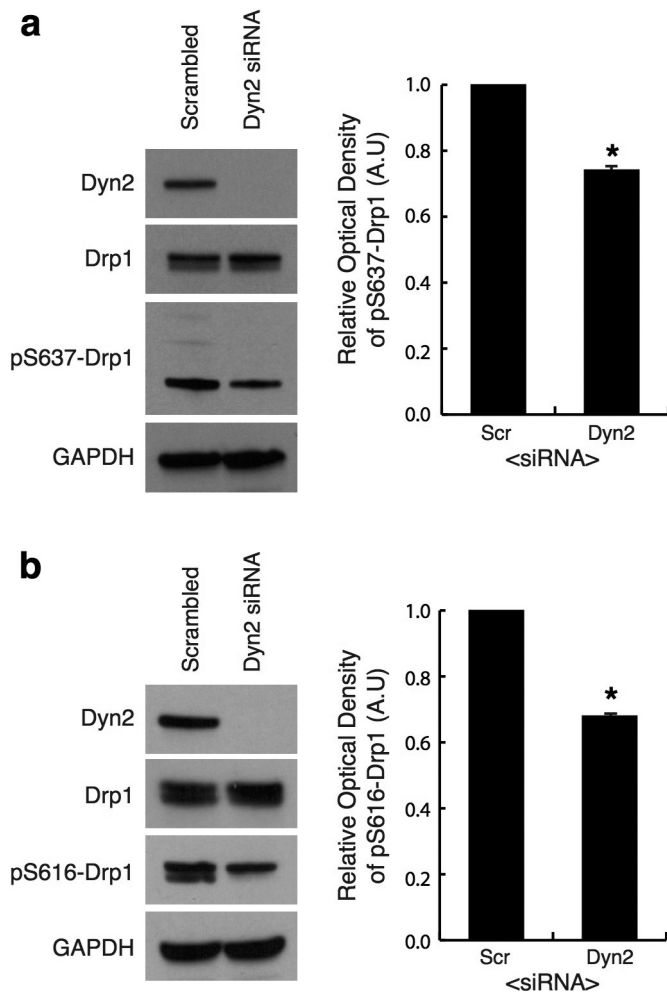
Extended Data Figure 1 | Dynamin-2 is required for mitochondrial division. **a**, Representative images of scrambled-, Drp1- or Dyn2-siRNA-transfected HeLa cells expressing mito-BFP. $n = 36$ cells for each siRNA treatment. Scale bars, $10\ \mu\text{m}$. **b**, Immunoblots with antibodies against Dyn2, Drp1, and GAPDH in cells treated with the indicated siRNA. **c**, The effect on mitochondrial morphology was quantified within a $230\ \mu\text{m}^2$ region of interest (ROI) for mean mitochondrial area per mitochondrion (left graph), and mean mitochondria per ROI (right graph).

As in COS-7 cells, Drp1- or Dyn2-depleted cells had larger mitochondria and fewer mitochondria per ROI when compared to control cells. Data were obtained from three biological replicate experiments for each of scrambled siRNA, Drp1 siRNA and Dyn2 siRNA treatments. Error bars represent s.e.m.; $*P < 0.01$ by ANOVA. **d**, Immunoblot analyses were performed on scrambled- and Dyn2-siRNA-treated cell lysates with antibodies against Dyn2, GAPDH, and mitochondrial fission (Drp1 and Mff) and fusion (Mfn2 and Opa1) machineries.

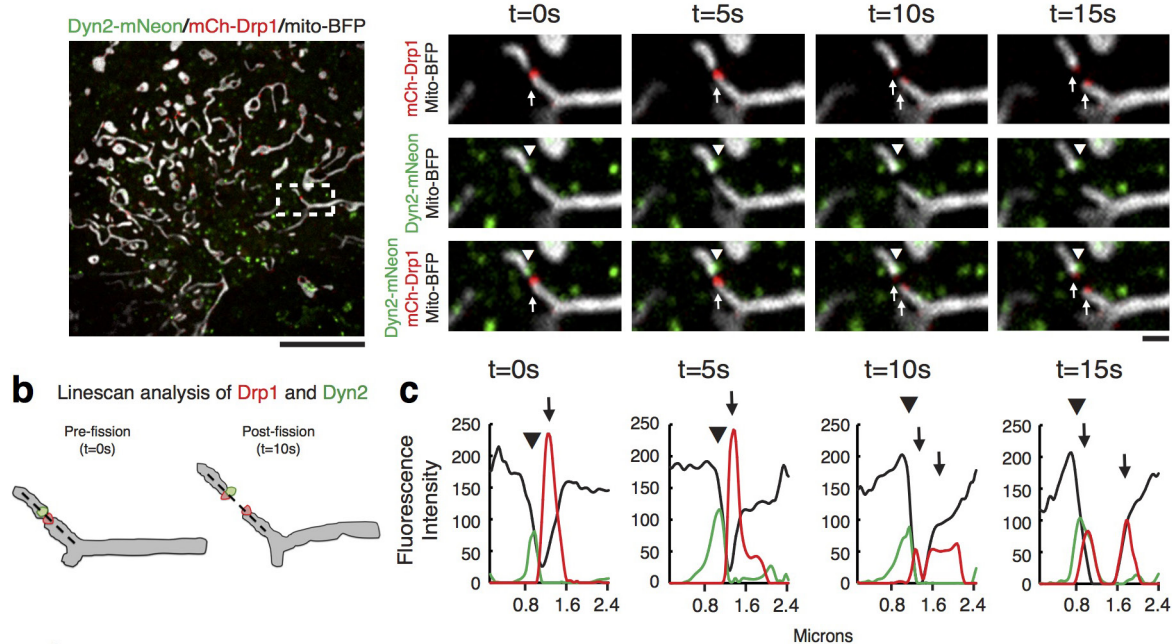
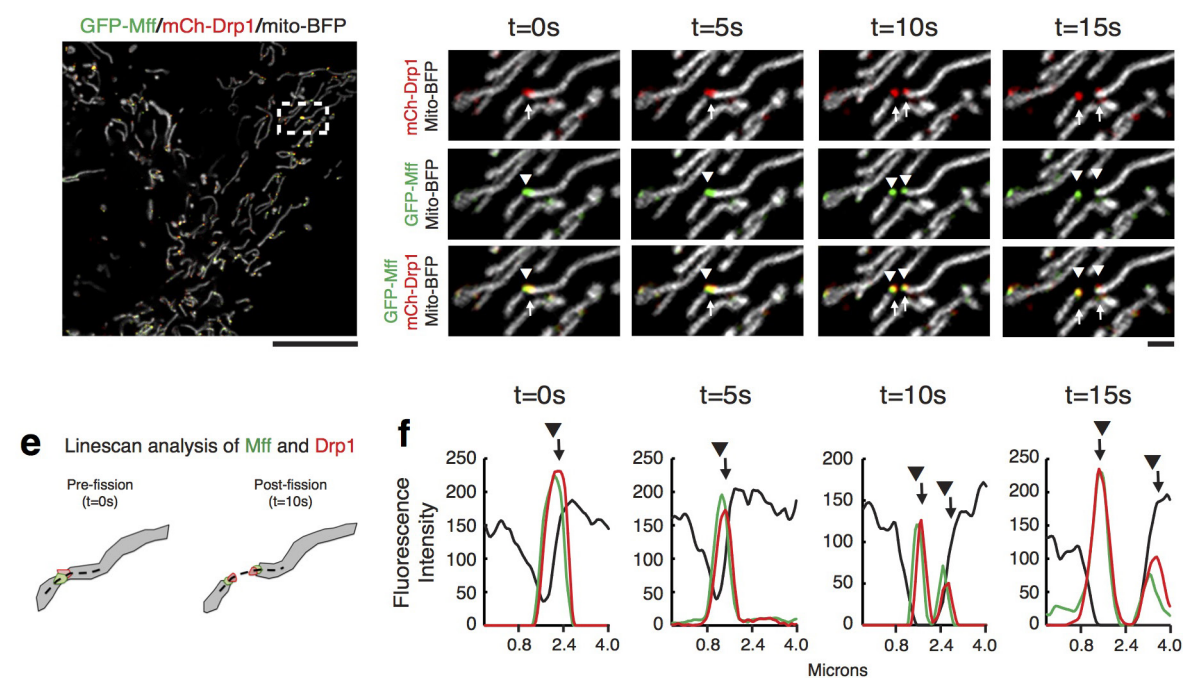
a COS-7 cells

Extended Data Figure 2 | Inhibition of clathrin-mediated endocytosis does not affect mitochondrial morphology. **a**, Representative images of TOM20 immunofluorescence in COS-7 cells transfected with scrambled control ($n = 50$), AP-2 ($n = 50$), or Dyn2 siRNA ($n = 52$). Scale bars = $10 \mu\text{m}$. **b**, Immunoblots with antibodies against AP-2, Dyn2, and GAPDH in siRNA-treated cells. **c**, The effect on mitochondrial morphology was quantitated within a $230 \mu\text{m}^2$ ROI for mean

mitochondrial area per mitochondrion (left) and mean mitochondria per ROI (right). Dyn2-depleted cells had larger mitochondria and less mitochondria per ROI compared to control cells, as in Fig. 1; however, AP-2-depleted cells displayed mitochondrial morphology that was qualitatively and quantitatively similar to control cells. These data were obtained from three biological replicate experiments. Error bars represent the s.e.m. $*P < 0.01$ by ANOVA.

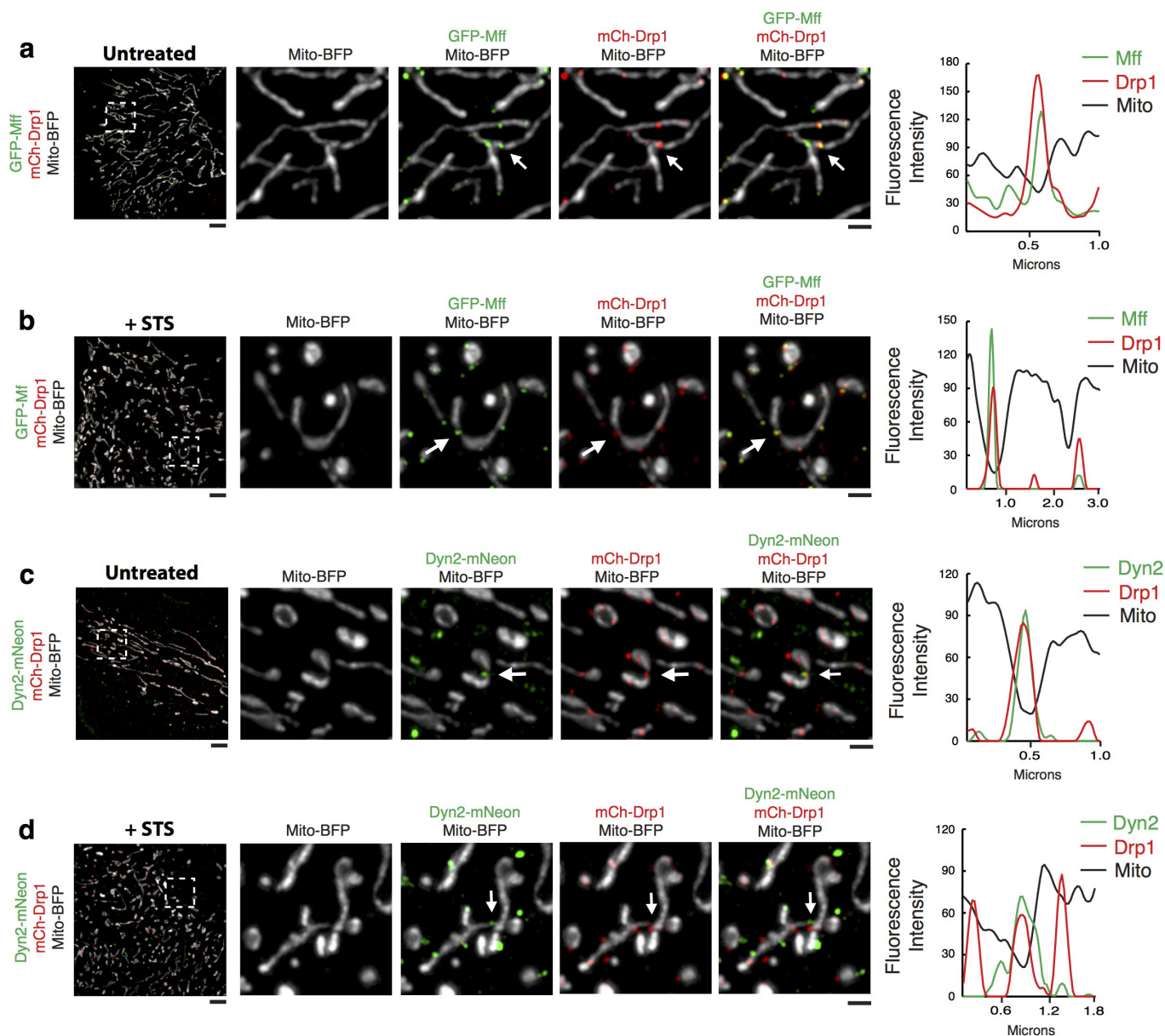


Extended Data Figure 3 | Dyn2 depletion does not phenocopy starvation-induced inhibition of Drp1. **a, b,** The phosphorylation status of Drp1 was evaluated by immunoblot on whole-cell lysates in scrambled- and Dyn2-siRNA-treated cells using antibodies against phospho-Ser637-Drp1 (**a**) and phospho-Ser616-Drp1 (**b**). Antibodies against Drp1 and Dyn2 were used to measure total Drp1 and Dyn2 levels, respectively, and anti-GAPDH was used as a loading control. The optical densities of phosphorylated-serine Drp1 were normalized to their corresponding GAPDH signal (graphs in **a, b**). The data represented in graphs in both **a** and **b** were obtained from three biological replicate experiments. Error bars represent the s.e.m. * $P < 0.01$ by paired t -test.

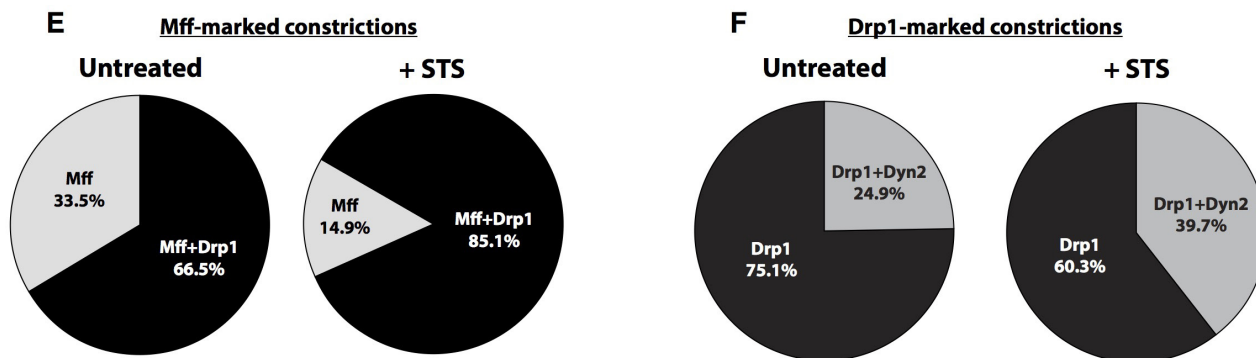
a COS-7 cells**d PtK1 cells**

Extended Data Figure 4 | Live-cell imaging of mitochondrial division machinery before, during and after division. **a**, Representative example of mitochondrial division (6 events from 108 cells) in COS-7 cells expressing mito-BFP (grey), mCherry-Drp1 (red), and Dyn2-mNeon (green) (Supplementary Video 2). Insets show the temporal and spatial dynamics of Drp1 (arrows) and Dyn2 (arrowheads) before and during mitochondrial division. **b**, A schematic identifying two temporal moments of interest with respect to Drp1, Dyn2 and mitochondrial dynamics—the frame before and the frame after division—with a dashed line that identifies the region that was analysed by line scan. **c**, Line-scan analysis of Drp1 and Dyn2 leading up to, and following, mitochondrial division.

d, Representative example mitochondrial division (27 events from 49 cells) in PtK1 cells expressing mito-BFP (grey), mCherry-Drp1 (red), and GFP-Mff (green). Insets show the temporal and spatial dynamics of Drp1 (arrows) and Mff (arrowheads) before and during mitochondrial division. **e**, Schematic identifying the two temporal moments of interest with respect to Mff, Drp1 and mitochondrial dynamics. **f**, Line-scan analysis of Mff and Drp1 leading up to and following mitochondrial division. Note, that the interaction between Drp1 and its adaptor, Mff, is maintained throughout the process of mitochondrial division; by contrast, Dyn2 associates with only one daughter mitochondrion. Scale bars for whole-cell panels and inset panels are 10 μm and 1 μm , respectively (**a**, **d**).

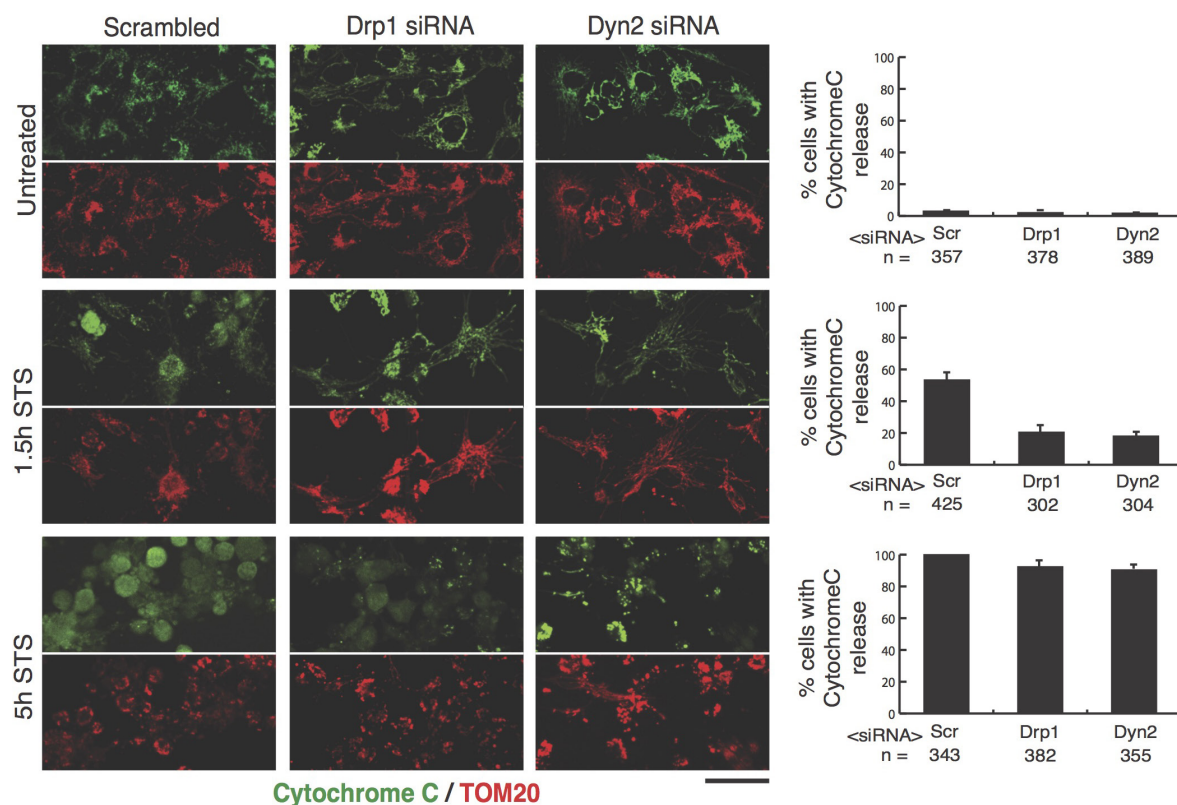
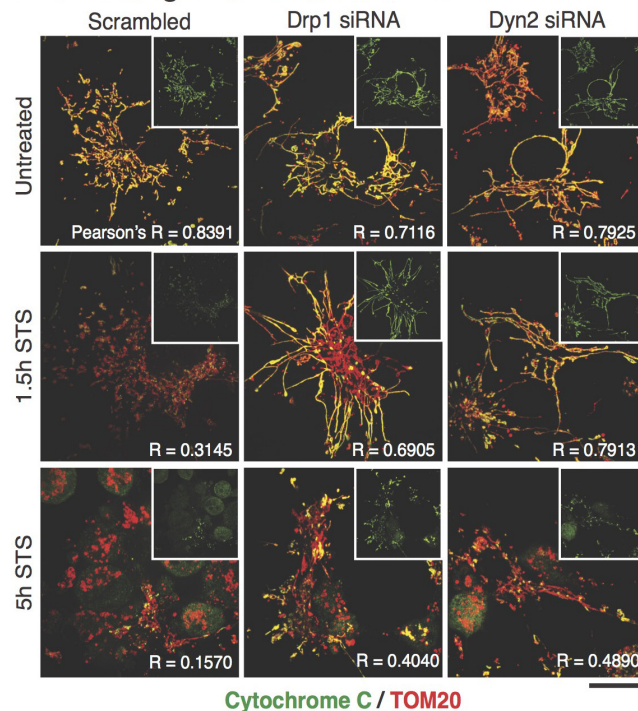


Mitochondria division factors at constrictions



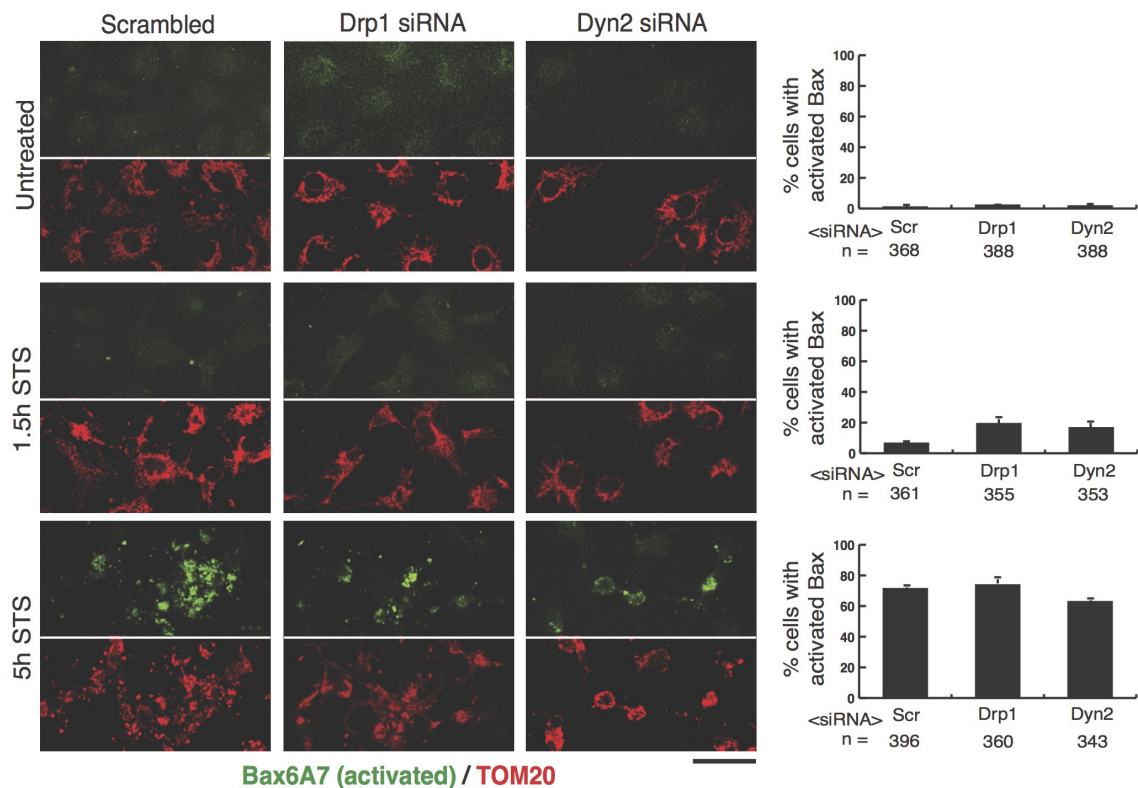
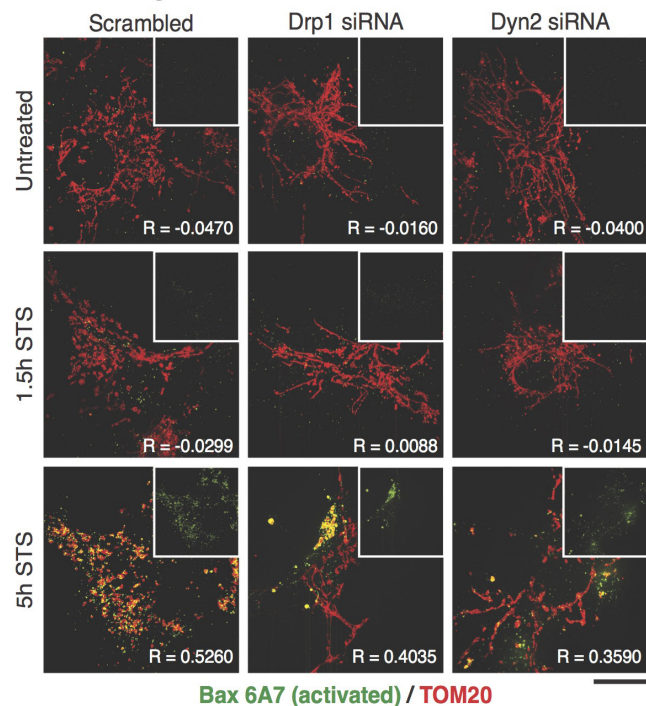
Extended Data Figure 5 | STS treatment stalls division factors at mitochondrial constrictions. a–d, Structure illumination microscopy was used to capture images of PtK1 cells expressing mCherry–Drp1 (red), mito–BFP (grey), and either GFP–Mff (a and b; green, $n = 10$ and 12 cells, respectively) or Dyn2–mNeon (c and d, green, $n = 13$ cells for each) that were untreated or treated with 1 μ M staurosporine (STS). The effect of STS treatment on division machinery localization was scored by line-scan analyses. Line-scan analysis verified the co-localization of Drp1 at Mff-marked constrictions (a, b) and Dyn2 at Drp1-marked constrictions (c, d). Under steady-state conditions, 66.5% of Mff-marked constrictions

co-labelled with Drp1 (336 out of 505 Mff-marked constrictions), whereas STS treatment increased the co-localization of Drp1 with Mff-marked constrictions to 85.1% (538 out of 632 Mff-marked constrictions). e, 24.9% of Drp1-marked constrictions were co-labelled with Dyn2 in untreated cells (128 out of 514 constrictions). The co-localization of Dyn2 to Drp1-marked constrictions increased to 39.7% following STS treatment (140 out of 353 Drp1-marked constrictions). f, STS treatment results in an increase of Drp1 at Mff-marked constrictions as well as an increase in Dyn2 localization to Drp1-marked constrictions. Scale bars for whole cell panels and the inset panel are 10 μ m and 1 μ m, respectively (a–d).

a 20x images of STS treated cells**b** 100x images of STS treated cells

Extended Data Figure 6 | Drp1 and Dyn2 depletion delays cytochrome *c* release from mitochondria after STS treatment. **a, b,** Scrambled-, Drp1-, and Dyn2-siRNA-treated cells were first treated with 75 μ M zVAD-fmk for 4 h, then either left untreated or treated with STS for 1.5 or 5 h. Cells were then fixed with 4% paraformaldehyde, permeabilized with 0.1% Triton X-100, and immuno-labelled for cytochrome *c* (green) and TOM20 (red). **a,** The percentage of cells displaying cytochrome *c* release was scored. *n* values are shown on each graph and depict the total

number of cells scored. Error bars represent s.e.m. **b,** Spatial resolution of the subcellular localization of cytochrome *c*. The main panel displays a merged image of cytochrome *c*/TOM20 with a single-channel image of cytochrome *c* (inset). The co-localization of cytochrome *c* with mitochondria was analysed using the 'coloc2' ImageJ plugin and the mean Pearson's *R* coefficient is displayed in the bottom-right corner of each image. *n* = 40 cells per condition. Data were obtained from three biological replicates. Scale bars, 50 μ m (**a**) or 10 μ m (**b**).

a 20x images of STS treated cells**b 100x images of STS treated cells**

Extended Data Figure 7 | Bax activation is accelerated in Drp1- and Dyn2-depleted cells following STS treatment. Scrambled-, Drp1-, and Dyn2-siRNA-treated cells were fixed as in Extended Data Fig. 6 and immuno-labelled with an antibody targeting activated Bax (Bax6A7; green), and TOM20 to stain mitochondria (red). **a**, The percentage of cells displaying Bax activation from the images taken (representative images shown on the left) was scored and plotted (right). *n* values represent the total number of cells scored. Error bars represent the s.e.m. **b**, Spatial

resolution of the subcellular localization of activated Bax. The main panel displays a merged image of activated Bax/TOM20, a single-channel image of activated Bax is shown in the inset. The co-localization of activated Bax with mitochondria was analysed using the 'coloc2' ImageJ plugin. The mean Pearson's *R* coefficient is displayed in the bottom-right corner of each image. *n* = 40 cells per condition. Data were obtained from three biological replicates. Scale bars, 50 μ m (**a**) or 10 μ m (**b**).

In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration

Keiichiro Suzuki^{1*}, Yuji Tsunekawa^{2*}, Reyna Hernandez-Benitez^{1,3*}, Jun Wu^{1,4*}, Jie Zhu^{5,6}, Eui-seok J. Kim⁷, Fumiyuki Hatanaka¹, Mako Yamamoto¹, Toshikazu Araoka^{1,4}, Zhe Li⁸, Masakazu Kurita¹, Tomoaki Hishida¹, Mo Li¹, Emi Aizawa¹, Shicheng Guo⁸, Song Chen⁸, April Goebel¹, Rupa Devi Soligalla¹, Jing Qu^{9,10}, Tingshuai Jiang^{6,11}, Xin Fu^{5,6}, Maryam Jafari⁶, Concepcion Rodriguez Esteban¹, W. Travis Berggren¹², Jeronimo Lajara⁴, Estrella Nuñez-Delicado⁴, Pedro Guillen^{4,13}, Josep M. Campistol¹⁴, Fumio Matsuzaki², Guang-Hui Liu^{10,15,16,17}, Pierre Magistretti³, Kun Zhang⁸, Edward M. Callaway⁷, Kang Zhang^{5,6,18,19} & Juan Carlos Izpisua Belmonte¹

Targeted genome editing via engineered nucleases is an exciting area of biomedical research and holds potential for clinical applications. Despite rapid advances in the field, *in vivo* targeted transgene integration is still infeasible because current tools are inefficient¹, especially for non-dividing cells, which compose most adult tissues. This poses a barrier for uncovering fundamental biological principles and developing treatments for a broad range of genetic disorders². Based on clustered regularly interspaced short palindromic repeat/Cas9 (CRISPR/Cas9)^{3,4} technology, here we devise a homology-independent targeted integration (HITI) strategy, which allows for robust DNA knock-in in both dividing and non-dividing cells *in vitro* and, more importantly, *in vivo* (for example, in neurons of postnatal mammals). As a proof of concept of its therapeutic potential, we demonstrate the efficacy of HITI in improving visual function using a rat model of the retinal degeneration condition retinitis pigmentosa. The HITI method presented here establishes new avenues for basic research and targeted gene therapies.

Site-specific transgene integration is typically achieved by the homology-directed repair (HDR) pathway including short-fragment homologous recombination (SFHR)^{5–7}, which is inefficient and not readily accessible to non-dividing cells⁸. By contrast, non-homologous end joining (NHEJ), the other major double strand break (DSB) repair pathway, is active in both proliferating and post-mitotic cells², and is generally more efficient than HDR in mammalian species⁹. Although mostly recognized as error-prone and used for generating targeted gene knockouts, studies have also demonstrated the intrinsic precision of NHEJ repair¹⁰, which was successfully harnessed for gene knock-ins^{11,12}. Regardless, however, NHEJ-mediated targeted transgene integration in post-mitotic cells has yet to be determined, especially *in vivo* in adult tissues such as the brain.

We aim to develop a robust NHEJ-based homology-independent strategy for targeted integration of transgenes in both dividing and non-dividing cells. First, we sought to improve upon existing NHEJ-based methods^{11,12} for more robust knock-in compared with HDR- and micro-homology-mediated end-joining (MMEJ)-based

methods¹³ using CRISPR/Cas9. To evaluate knock-in efficiencies we generated a GFP-correction HEK293 line (Fig. 1a). The absolute knock-in efficiencies via HDR, MMEJ-mediated targeted integration (precise integration into target chromosome (PITCh))¹³, or NHEJ-mediated targeted integration (designated herein as homology-independent targeted integration, or HITI) (Extended Data Fig. 1a), were presented as percentages of GFP⁺ or mCherry⁺ cells (Fig. 1a, b). We observed little to no knock-in events when using genome cut only (IRESmCherry-0c) and donor DNA cut only (IRESmCherry-MC-scramble) control donors (Fig. 1a, b and Extended Data Fig. 1b, c). Notably, we observed higher knock-in efficiency with HITI donors (IRESmCherry-1c, -2c and -MC; see below for definitions) than with an HDR donor (truncated GFP (tGFP) and IRESmCherry-HDR-0c), a PITCh donor (IRESmCherry-MH) or a HITI donor with homology arms (IRESmCherry-HDR-2c). Consistent with previous observations, inserted DNA devoid of bacterial backbone (IRESmCherry-2c and IRESmCherry-MC) resulted in less pronounced transgene silencing than DNA carrying bacterial sequences (IRESmCherry-1c) (Extended Data Fig. 1d–f)^{14,15}. Treatment with the NHEJ inhibitor NU7026 significantly decreased HITI efficiency, confirming the dependence of HITI on the NHEJ repair machinery (Extended Data Fig. 1g).

HITI is expected to occur more frequently in the forward than the reverse direction as an intact guide RNA (gRNA) target sequence remains in the latter, which is subjected to additional Cas9 cutting until forward transgene insertion or insertions and deletions (indels) occur that prevent further gRNA binding (Extended Data Fig. 1a). Indeed, we only found 1 in 48 mCherry[–] single-cell-derived clones showed reverse integration (Extended Data Fig. 2a). Notably, the majority of forward knock-in did not exhibit indels (Fig. 1b and Extended Data Fig. 2b–g). The GFP-correction HEK293 line contains five copies of the target site per cell. Next we sequenced all the target sites of mCherry⁺ and mCherry[–] single-cell clones (Extended Data Fig. 2h). Among 13 mCherry⁺ single-cell clones analysed, we observed 1–3 knock-in events per clone and the rest of the genomic targets all contained indels. By contrast, 22 of 24 mCherry[–] single-cell clones showed intact target

¹Gene Expression Laboratory, Salk Institute for Biological Studies, 10010 N. Torrey Pines Rd, La Jolla, California 92037, USA. ²Laboratory for Cell Asymmetry, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan. ³4700 King Abdullah University of Science and Technology (KAUST) Thuwal 23955-6900, Saudi Arabia. ⁴Universidad Católica San Antonio de Murcia (UCAM) Campus de los Jerónimos, no. 135 Guadalupe 30107, Murcia, Spain. ⁵Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China. ⁶Shiley Eye Institute, Institute for Genomic Medicine, Institute of Engineering in Medicine, University of California, San Diego, 9500 Gilman Drive #0946, La Jolla, California 92023, USA. ⁷Systems Neurobiology Laboratory, Salk Institute for Biological Studies, 10010 N. Torrey Pines Rd., La Jolla, California 92037, USA. ⁸Bioengineering, University of California, San Diego, 9500 Gilman Drive, MC0412, La Jolla, California 92093-0412, USA. ⁹State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. ¹⁰University of Chinese Academy of Sciences, Beijing 100049, China. ¹¹Guangzhou EliteHealth Biological Pharmaceutical Technology Company Ltd, Guangzhou 510005, China. ¹²Salk Institute for Biological Studies, 10010 N. Torrey Pines Rd, La Jolla, California 92037, USA. ¹³Fundación Dr. Pedro Guillen, Investigación Biomedica de Clínica CEMTRO, Avenida Ventisquero de la Condesa, 42, 28035 Madrid, Spain. ¹⁴Hospital Clinic, University of Barcelona, IDIBAPS, 08036 Barcelona, Spain. ¹⁵National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. ¹⁶Key Laboratory of Regenerative Medicine of Ministry of Education, Institute of Aging and Regenerative Medicine, Jinan University, Guangzhou 510632, China. ¹⁷Beijing Institute for Brain Disorders, Beijing 100069, China. ¹⁸Molecular Medicine Research Center, West China Hospital, Sichuan University, Chengdu 610041, China. ¹⁹Veterans Administration Healthcare System, San Diego, California 92093, USA.

*These authors contributed equally to this work.

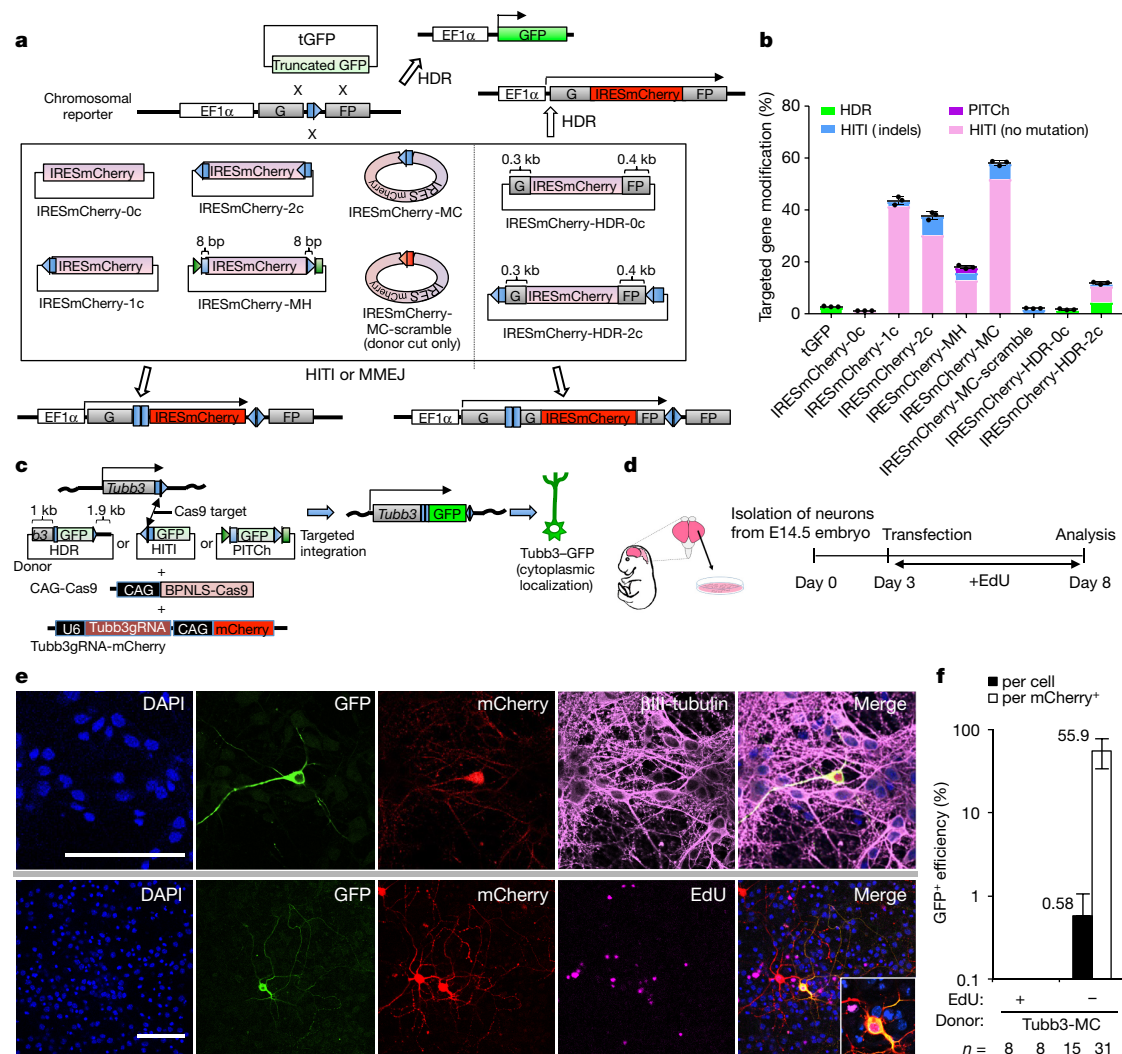


Figure 1 | HITI-mediated *in vitro* genome editing. **a**, Schematic of gene targeting by HDR, PITCh, or HITI in the GFP-correction HEK293 line. Blue pentagon, Cas9/gRNA target sequence. Black line within blue pentagon, Cas9 cleavage site. HITI donors: IRESmCherry-0c, IRESmCherry-1c, IRESmCherry-2c or IRESmCherry-MC. HDR donors: tGFP and IRESmCherry-HDR-0c. A PITCh-donor: IRESmCherry-MH. The green square and triangle, CRISPR/Cas9 target sites to create 8-bp microhomology at both ends of the IRESmCherry cassette. HDR and HITI dual donors: IRESmCherry-HDR-2c. Genome cut-only control donor: IRESmCherry-0c. Donor cut-only control donor: IRESmCherry-MC-scramble. Red pentagon, Scramble-gRNA target sequence. **b**, Gene targeting efficiency by HDR, PITCh, or HITI (with or without indels

at junction sites). Results were obtained from three replicate wells and presented as mean \pm s.d. The input data points were shown as black dots. **c**, Schematic of targeted GFP knock-in by HDR, PITCh, or HITI in cultured primary neurons. **d**, Experimental scheme for GFP knock-in in cultured primary neurons. **e**, Representative immunofluorescence images of neurons transfected with Tubb3-GFP HITI constructs: BPNLS-Cas9, Tubb3gRNA-mCherry, and Tubb3-MC plasmids. Inset, higher magnification image. Scale bar, 100 μ m. **f**, Targeted GFP knock-in efficiencies indicated by percentage of GFP⁺ cells among transfected cells (mCherry⁺) or all cells (DAPI⁺) in EdU⁺ or EdU⁻ neurons. *n*, technical replicates. Results were presented as mean \pm s.d. For source data, see Supplementary Table 1.

sequences. The remaining two mCherry⁻ clones showed either indels or reverse integration at all target sites.

To further enhance Cas9 activity and HITI editing, we tested fusing Cas9 to different nuclear localization signals (NLS) and found the bipartite SV40NLS or BPNLS¹⁶ was superior to SV40NLS⁴ in Cas9 nuclear targeting and genome editing (Extended Data Fig. 3).

Next we tested HITI in non-dividing cells *in vitro*. To this end we transfected cultured mouse primary neurons with HITI constructs designed to insert a GFP cassette downstream of the *Tubb3* gene, which would result in the expression of a TUBB3-GFP fusion protein localized to the cytoplasm¹⁷. We used EdU to label proliferating cells. Five days post-transfection we observed many neurons with GFP signal co-localized with β III-tubulin/Tuj1 (Fig. 1c–e) and were EdU-negative, indicating successful HITI-mediated GFP knock-in to the *Tubb3* locus in post-mitotic neurons (Fig. 1f and Extended Data Fig. 4a, b). The percentage of GFP⁺ cells was 0.58% of total cell population (GFP⁺/DAPI⁺,

absolute efficiency) and 55.9% of transfected cells (GFP⁺/mCherry⁺, relative efficiency), respectively.

We compared relative knock-in efficiencies in transfected neurons with an HDR (Tubb3-HDR), a PITCh (Tubb3-MH), and four different HITI donor plasmids (1-cut (Tubb3-1c), 2-cut (Tubb3-2c), 2-cut No-polyA (Tubb3-2cd), and minicircle (Tubb3-MC)) (Extended Data Fig. 4c). We observed little to no knock-ins with donor DNA cut-only control (Tubb3-MC-scramble), HDR (Tubb3-HDR), and PITCh (Tubb3-MH) vectors. By contrast, all HITI vectors allowed efficient GFP knock-in in non-dividing primary neurons. For the 2-cut No-polyA and minicircle donors, the GFP signal was almost exclusively found in the cytoplasm, while 1-cut and 2-cut donors often resulted in mis-localized GFP signals (Extended Data Fig. 4d, e), probably due to the presence of bacterial backbone and/or polyA sequences. Consistent with our observations in HEK293 cells, the majority of knock-in neurons (GFP⁺) did not show indels and we observed a

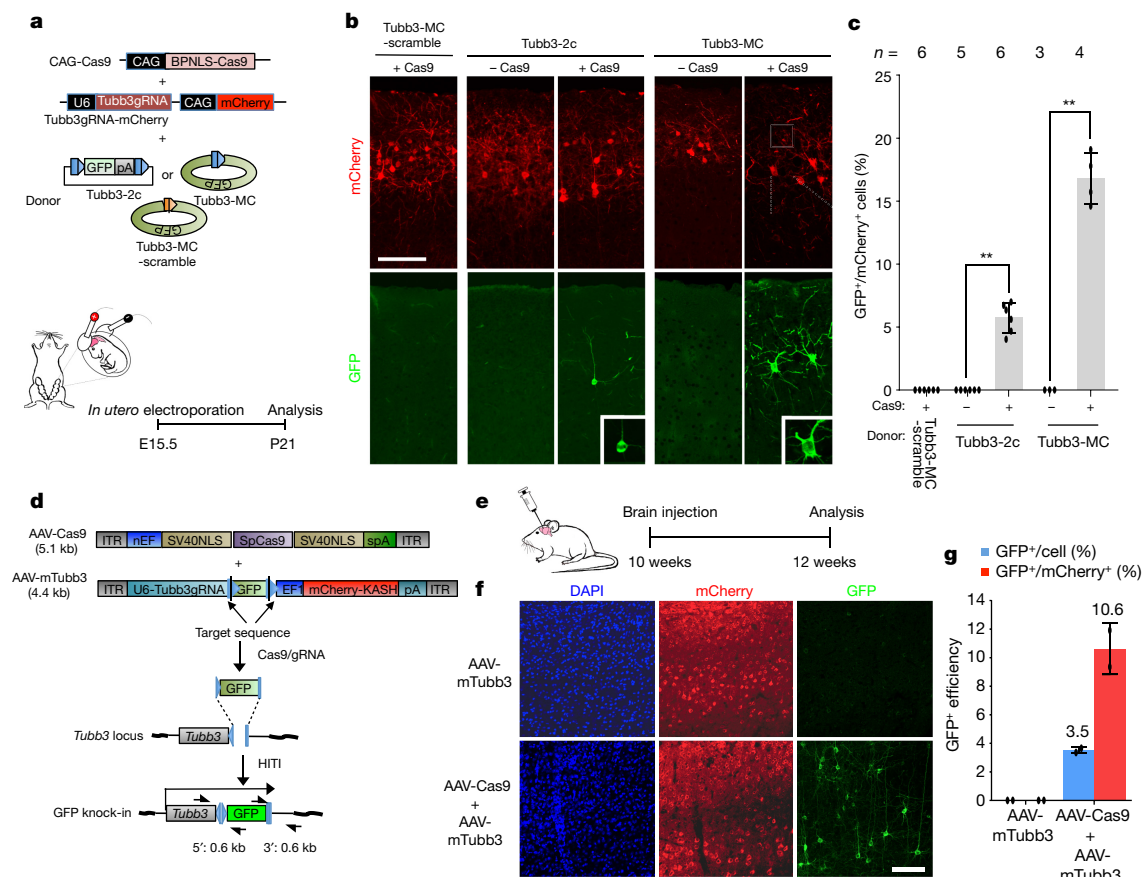


Figure 2 | HITI-mediated *in vivo* genome editing in neurons.

a, Schematic of targeted GFP knock-in by HITI in fetal brain via *in utero* electroporation. **b**, Representative immunofluorescence images of neurons showing correct GFP knock-in at the *Tubb3* locus. Insets, higher magnification images. Scale bar, 100 μ m. mCherry, transfected cells. **c**, Relative knock-in efficiency measured by the percentage of GFP⁺ cells among mCherry⁺ cells. *n*, number of pups obtained from two pregnant mice. Results were presented as mean \pm s.d. The input data points were shown as black dots. ***P* < 0.01, unpaired Student's *t*-test.

better knock-in efficiency with BPNLS than SV40NLS (Extended Data Fig. 4f–i and Extended Data Fig. 5a). Furthermore, HITI also allowed knock-in of GFP to the *TUBB3* locus in human embryonic stem cell (ES cell)-derived pan neurons (Extended Data Fig. 5b–e).

We next tested HITI *in vivo* using live rodents. We first delivered Tubb3–GFP HITI constructs to E15.5 mouse brain via *in utero* electroporation¹⁸. Three weeks after birth, we observed that up to 16.8% of electroporated cells (GFP⁺/mCherry⁺, relative efficiency) showed cytoplasmic GFP expression (Fig. 2a–c and Extended Data Fig. 5f). Owing to low efficiency of DNA delivery via *in utero* electroporation (around 1%), the absolute knock-in efficiency (GFP⁺/DAPI⁺) to the somatosensory cortex was less than 0.2%. To demonstrate the efficacy of HITI in post-mitotic cells *in vivo*, we generated inducible Tubb3–GFP HITI targeting constructs where Cre-dependent Cas9 expression is under the control of tamoxifen (TAM). Following *in utero* electroporation of E15.5 fetal brain we induced Cas9 expression via TAM treatment at postnatal day (P)10 and P11, at which stage most, if not all, dividing neural progenitors have differentiated into post-mitotic neurons¹⁹ (Extended Data Fig. 5g). Our results revealed efficient GFP knock-in with the HITI donor and minimal knock-in with the HDR donor (Tubb3-HDR) (Extended Data Fig. 5h, i). Similar results were obtained for other tissues, such as muscle and kidney, where non-viral delivery was used for HITI constructs that target GFP to the *Rosa26* locus downstream of the CAG promoter in Ai14 mice²⁰ (Extended Data Fig. 6). Together, these results demonstrate the utility of HITI for

d, Schematic of HITI-AAV vectors for knock-in of GFP to the downstream of *Tubb3* gene. Black half-arrows, PCR primers. **e**, Schematic of *in vivo* HITI via local AAV injections in adult mouse brain. **f**, Representative immunofluorescence images of neurons in HITI-AAV injected brain sections. Scale bar, 100 μ m. **g**, Relative and absolute knock-in efficiencies measured by the percentage of GFP⁺ cells among mCherry⁺ cells or all DAPI⁺ cells, respectively. Results were obtained from two animals and presented as mean \pm s.d. The input data points were shown as black dots. For source data, see Supplementary Table 2.

transgene knock-in in a variety of somatic tissues, including non-dividing cells, *in vivo*.

For *in vivo* applications, adeno-associated virus (AAV) vectors have been the delivery method of choice and have shown evidence of efficacy in humans^{21–23}. Thus, to further improve *in vivo* HITI efficiency and utility further, we sub-cloned HITI constructs into two AAV vectors and packaged both AAVs with serotype 8 or 9, which showed high infection capability for many organs and therapeutic safety^{24,25} (Fig. 2d). Efficacy of HITI-AAVs (AAV-Cas9 and AAV-mTubb3) was first tested by co-infecting cultured primary neurons. Infected neurons subsequently expressed GFP overlapped with cytoplasmic β III-tubulin staining, and the percentage of GFP⁺ cells was 5.1% (absolute efficiency) of the total cell population (Extended Data Fig. 7a–c). The fidelity of GFP knock-in was confirmed by PCR and sequencing (Extended Data Fig. 7d, e).

We next injected the HITI-AAVs directly to the visual cortex of adult mouse brain. Two weeks later, brain sections were stained with a GFP antibody (Fig. 2e). In contrast to control (AAV-mTubb3 alone), where minimal GFP signals were detected, we found many neurons with cytoplasmic GFP signals in the experimental group (Fig. 2f). Our results indicate that the percentage of GFP⁺ cells in the adult mouse brain was 3.5% (absolute efficiency) of total cells counted and 10.6% (relative efficiency) of infected cells (mCherry⁺), respectively (Fig. 2g). We also generated HITI-AAVs for GFP knock-in in Ai14 mice and found local HITI-AAVs delivery through intramuscular injection in

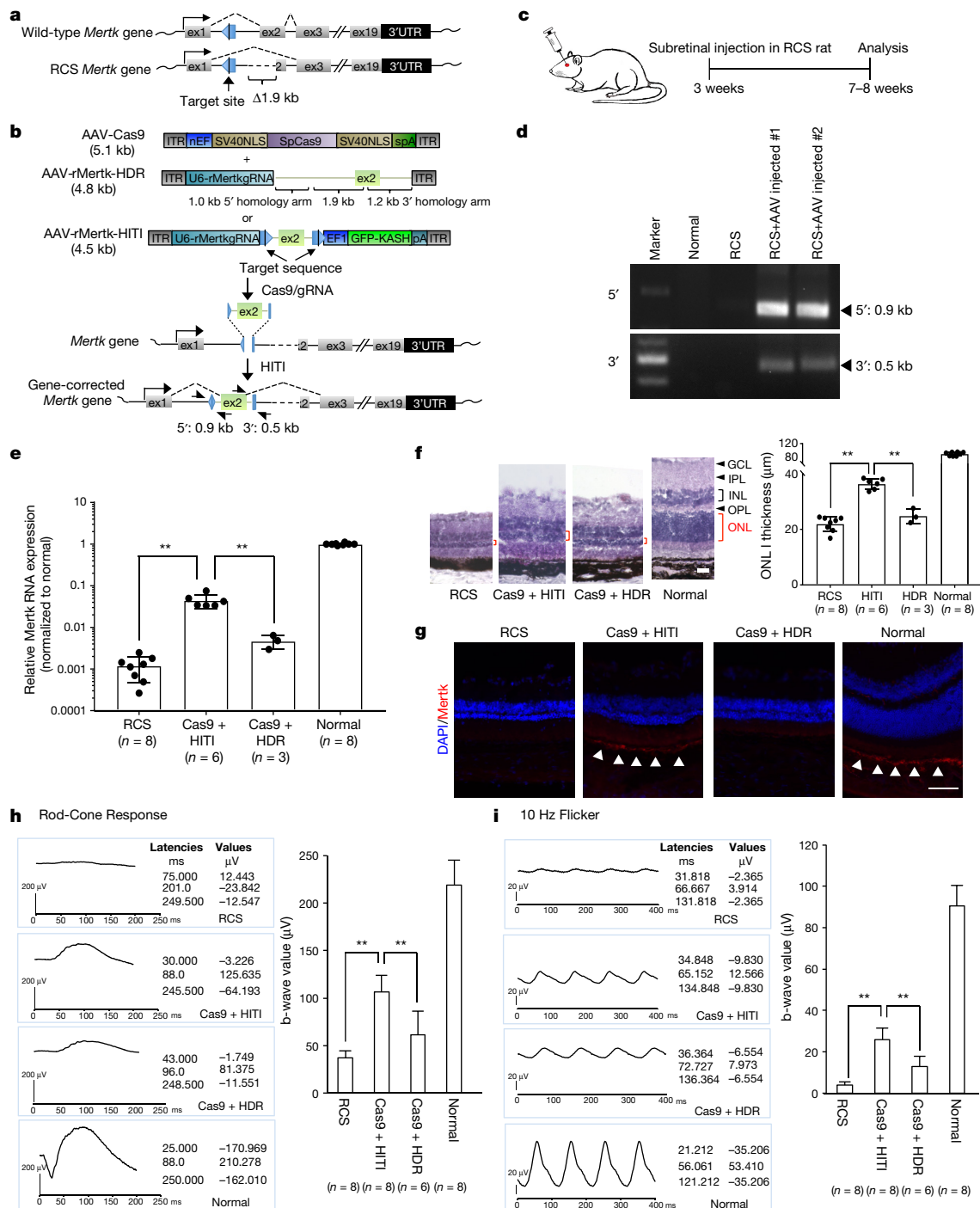


Figure 3 | HITI-mediated gene correction of a rat model of retinitis pigmentosa. **a**, Schematic representation of the *Mertk* gene locus in both wild-type and RCS rats. Blue pentagon, Cas9/gRNA target sequence. Black line within blue pentagon, Cas9 cleavage site. **b**, Schematic of *Mertk* gene correction AAV vectors. Black half-arrows, PCR primers. **c**, Experimental scheme for *Mertk* gene correction in RCS rats. **d**, Validation of correct gene knock-in by genomic PCR. **e**, Relative *Mertk* mRNA expression. *n*, number of animals. Results were presented as mean \pm s.d. The input data points were shown as black dots. ** $P < 0.01$, unpaired Student's *t*-test. **f**, Retinal morphology and outer nuclear layer (ONL) thickness. Red bracket, ONL. GCL, ganglion cell layer; IPL, inner plexiform layer; INL, inner

adult Ai14 mice also resulted in GFP knock-in at the *Rosa26* locus in skeletal muscle (Extended Data Fig. 8a–c).

To explore the possibility of using HITI for gene replacement therapy, we used the Royal College of Surgeons (RCS) rat, a well-established animal model for retinitis pigmentosa (an inherited retinal

degeneration condition that causes blindness in humans²⁶) resulting from a homozygous 1.9-kb deletion from intron 1 to exon 2 in the *Mertk* gene²⁷ (Fig. 3a). Morphological changes in the degenerating photoreceptor outer nuclear layer (ONL) appear as early as P16 in RCS rats. To try to restore MERTK function in the eye, we generated a HITI-AAV

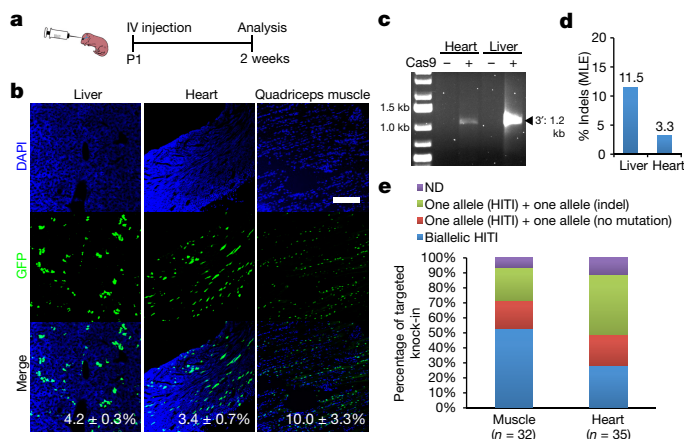


Figure 4 | AAV-mediated systemic HITI *in vivo*. **a**, Schematic of *in vivo* targeted GFP-NLS knock-in by HITI via intravenous (IV) AAV injections to Ai14 mice. **b**, Representative immunofluorescence images of GFP expression in the liver, heart, and quadriceps muscle after intravenous injection of HITI-AAVs (serotype 9). The absolute efficiencies of GFP knock-in per cell (liver and heart) and per nucleus (quadriceps muscle) were shown. Results were obtained from five sections per sample and presented as mean \pm s.d. Scale bar, 200 μ m. **c**, Validation of correct gene knock-in in heart and liver by genomic PCR. **d**, Frequencies of indel per target sequence of liver and heart by deep sequencing. **e**, Single-cell genotyping of GFP⁺ cells from HITI-AAVs injected Ai14 mouse muscle and heart. *n*, number of the analysed single cells. ND, not determined. For full gel image, see Supplementary Fig. 1. For source data, see Supplementary Table 4.

vector for inserting a copy of *Mertk* exon 2 into intron 1 (AAV-rMertk-HITI) and an HDR-AAV vector (AAV-rMertk-HDR) as a control (Fig. 3b). The AAVs were injected in the subretinal space of RCS rats 3 weeks after birth and analyses were performed between 7–8 weeks (Fig. 3c). DNA analysis revealed correct knock-in in eyes injected with HITI-AAVs (Fig. 3d and Extended Data Fig. 8d). In addition, HITI-AAVs injection led to statistically significant increases in *Mertk* mRNA expression levels (absolute efficiency: 4.5% of wild-type level) and better preservation of the ONL thickness compared with non-injected and HDR-AAV-injected controls (Fig. 3e, f). MERTK protein was also observed in eyes injected with HITI-AAVs, but not HDR-AAVs (Fig. 3g). To determine the effect of HITI treatment on the physiological function of the retina, electroretinography (ERG) responses were tested to measure the electrical activity of rods and cones (10 Hz flicker). The ERG tests were performed 4 weeks after the injection (P50). All eyes treated with HITI-AAVs exhibited significantly improved ERG b-wave responses (Fig. 3h). Similarly, 10 Hz flicker value, which measures cone response, was significantly improved and was more than fourfold higher than seen for the untreated eyes (Fig. 3i). It should be noted, however, that although tests demonstrated improved visual responses after subretinal injection of HITI-AAVs to 3-week-old RCS rats, the rescue was only partial and not enough to completely restore vision. Thus, when considering a putative path forward for treating this disease in the clinic, early intervention times (before overt retinal degeneration is seen) would be required.

We also tested systemic delivery of HITI-AAVs (AAV-Cas9 and AAV-Ai14-HITI) into P1 Ai14 mice via intravenous injection (Fig. 4a). Two weeks post-injections, we observed 4.2%, 3.4% and 10.0% GFP⁺ cells of total nuclei counted in the liver, heart and quadriceps muscle, respectively (GFP⁺/DAPI⁺, absolute efficiency) (Fig. 4b). GFP signals were also detected in a broad range of other organs and tissues (Extended Data Fig. 9a). We also compared *in vivo* gene knock-in efficiencies between HITI and HDR via systemic delivery and found that knock-in efficiency with HITI was significantly higher than with HDR in both liver and heart tissues (Extended Data Fig. 9b, c). Genomic PCR and DNA sequencing analyses of whole, unsorted liver and heart samples confirmed the correct gene knock-ins (Fig. 4c and

Extended Data Figs 8a and 9d). We also found that indel frequencies, as determined by deep sequencing of unsorted samples, were comparable to knock-in efficiencies, similar to our observation in HEK293 cells (Fig. 4b, d and Extended Data Fig. 2h). Next we performed single-cell genotyping using GFP⁺ cells sorted from muscle and heart tissues (Fig. 4e). Our results demonstrated high on-target specificity of HITI (90–95%). Notably, among all analysed cells, 30–50% showed biallelic transgene integration. In addition, we observed preferential knock-in in adult liver after tail vein injection of AAVs, with an absolute efficiency of approximately 5% of total cells counted (Extended Data Fig. 10a–d). To study off-target effects of HITI following systemic *in vivo* delivery, we examined mutation rates for on-target sites, and the 12 highest ranked predicted genomic off-target sites using liver tissue. Next-generation sequencing analysis revealed minimal indel frequency at the examined off-target sites (Extended Data Fig. 10e).

The ability to use HITI for *in vivo* targeted transgene insertion into post-mitotic neurons is unprecedented and will help advance basic and translational neuroscience research. For example, HITI-mediated insertion of optogenetic activators into downstream of a specified gene locus may help gain cell-type-specific control over neuronal activities²⁸. HITI may also allow, for instance, the generation of knock-in reporters for tracing cells in live animals. This will be particularly useful for animal models in which transgenic tools are limited (for example, non-human primates)²⁹. With further improvement to efficiency, HITI also holds great promise for *in vivo* targeted gene-replacement therapy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 February; accepted 27 October 2016.

Published online 16 November 2016.

- Naldini, L. Gene therapy returns to centre stage. *Nature* **526**, 351–360 (2015).
- Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).
- Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
- Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Lombardo, A. et al. Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat. Methods* **8**, 861–869 (2011).
- Genovese, P. et al. Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **510**, 235–240 (2014).
- Luchetti, A., Maligni, A. & Sangiuolo, F. Small fragment homologous replacement (SFHR): sequence-specific modification of genomic DNA in eukaryotic cells by small DNA fragments. *Methods Mol. Biol.* **1114**, 85–101 (2014).
- Orthwein, A. et al. A mechanism for the suppression of homologous recombination in G1 cells. *Nature* **528**, 422–426 (2015).
- Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).
- Bétermier, M., Bertrand, P. & Lopez, B. S. Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet.* **10**, e1004086 (2014).
- Maresca, M., Lin, V. G., Guo, N. & Yang, Y. Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res.* **23**, 539–546 (2013).
- Auer, T. O., Duroure, K., De Cian, A., Concordet, J.-P. & Del Bene, F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* **24**, 142–153 (2014).
- Nakade, S. et al. Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* **5**, 5560 (2014).
- Chen, Z.-Y. et al. Linear DNAs concatamerize *in vivo* and result in sustained transgene expression in mouse liver. *Mol. Ther.* **3**, 403–410 (2001).
- Chen, Z.-Y., He, C.-Y., Ehrhardt, A. & Kay, M. A. Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression *in vivo*. *Mol. Ther.* **8**, 495–500 (2003).
- Wu, J., Corbett, A. H. & Berland, K. M. The intracellular mobility of nuclear import receptors and NLS cargoes. *Biophys. J.* **96**, 3840–3849 (2009).
- Kameda, Y., Kameya, T. & Frankfurter, A. Immunohistochemical localization of a neuron-specific β -tubulin isotype in the developing chicken ultimobranchial glands. *Brain Res.* **628**, 121–127 (1993).
- Tabata, H. & Nakajima, K. Efficient *in utero* gene transfer system to the developing mouse brain using electroporation: visualization of neuronal migration in the developing cortex. *Neuroscience* **103**, 865–872 (2001).

19. Mizutani, K. & Saito, T. Progenitors resume generating neurons after temporary inhibition of neurogenesis by Notch activation in the mammalian cerebral cortex. *Development* **132**, 1295–1304 (2005).
20. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
21. Samulski, R. J. & Muzyczka, N. AAV-mediated gene therapy for research and therapeutic purposes. *Annu. Rev. Virol.* **1**, 427–451 (2014).
22. Bainbridge, J. W. B. *et al.* Effect of gene therapy on visual function in Leber's congenital amaurosis. *N. Engl. J. Med.* **358**, 2231–2239 (2008).
23. Maguire, A. M. *et al.* Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N. Engl. J. Med.* **358**, 2240–2248 (2008).
24. Nathwani, A. C. *et al.* Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N. Engl. J. Med.* **371**, 1994–2004 (2014).
25. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV serotypes 1–9 mediated gene expression and tropism in mice after systemic injection. *Mol. Ther.* **16**, 1073–1080 (2008).
26. Luo, J. *et al.* Human retinal progenitor cell transplantation preserves vision. *J. Biol. Chem.* **289**, 6362–6371 (2014).
27. D'Cruz, P. M. *et al.* Mutation of the receptor tyrosine kinase gene *Mertk* in the retinal dystrophic RCS rat. *Hum. Mol. Genet.* **9**, 645–651 (2000).
28. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
29. Izpisua Belmonte, J. C. *et al.* Brains, genes, and primates. *Neuron* **86**, 617–631 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to M. Kay, Z. Y. Chen, G. Lemke and P. G. Burrola for sharing experimental materials; J. Naughton, L. Lisowski and J. Marlett for AAV production; C. Fine, J. Olvera, E. O'Connor and K. E. Marquez for cell sorting; D. Okamura and M. Jacobs for mouse surgery and histology processing; D. Skowronska-Krawczyk for rat experiments; N. V. Gohad, T. Whitfield, I. M. Verma, J. Ogawa, T. Hara, U. Manor and J. Santini for imaging; L. Greg, Y. S. Kida and F. Osakada for valuable discussions; D. O'Keefe for proofreading the manuscript and M. Schwarz for administrative help. Core Facilities were utilized at the Salk Institute (support from: NIH-NCI CCSG: P30 014195, NINDS R24NS092943, and NEI P30 EY019005) and UCSD Neuroscience core grant P30 NS047101. R.H.B. was supported by a CONACYT fellowship of Mexico. J.Z. and T.J. were supported by 973 Program (2013CB967504, 2015CB964600) and 863 Program (2014AA021604). T.H. was partially supported by a Nomis Foundation Fellowship. E.J.K. is a

Biogen-IDEF Fellow of the Life Science Research Foundation. M.Y. was partially supported by the Salk Women & Science Special Award. X.F. was supported by NSFC (No. 81601872). G.H.L. and J.Q. were supported by the National Basic Research Program of China (973 Program; 2015CB964800, 2014CB910503, 2013CB967504), National Natural Science Foundation of China (81625009, 81371342, 81271266), the National High Technology Research and Development Program of China (2015AA020307, 2014AA021604), and Program of Beijing Municipal Science and Technology Commission (Z151100003915072). F.M. was supported by RIKEN funding for Development and Regeneration. Ku.Z. was supported by NIH grant R01HL123755. P.J.M. and J.C.I.B. were supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under award no. OSR-2015-CRG4-2631. Work in the laboratory of J.C.I.B. was supported by The Leona M. and Harry B. Helmsley Charitable Trust (2012-PG-MED002), the G. Harold and Leila Y. Mathers Charitable Foundation, NIH (R01HL123755), The McKnight Foundation, The Moxie Foundation, Fundacion Dr. Pedro Guillen and Universidad Católica San Antonio de Murcia (UCAM).

Author Contributions K.S., Y.T., R.H.B., J.W. and J.C.I.B. conceived the project and designed experiments. K.S., Y.T. and E.A. constructed plasmids. K.S., R.H.B., M.Y. and M.L. generated minicircle DNA vectors. K.S., Y.T., R.H.B., M.L., E.A., A.G. and R.D.S. performed work on HEK293 cells. K.S., Y.T., R.H.B. and E.A. performed bisulfite sequencing. K.S., Y.T. and R.H.B. measured intracellular localization of dCas9. K.S. and Y.T. performed the Surveyor assay. R.H.B. performed work on primary neurons. K.S., Y.T., R.H.B., E.A. and A.G. performed work on human ES-cell-derived pan neurons. Y.T. and F.M. performed work on *in utero* electroporation. J.Z., T.J., X.F., M.J. and Ka.Z. performed work on RCS rats. E.J.K. and E.M.C. performed work on adult mouse brain. F.H., T.A., M.K. and T.H. performed *in vivo* mouse electroporation. F.H., M.Y. and T.A. performed AAV IV and IM injection in neonatal or adult mice. Z.L., S.G., S.C. and Ku.Z. performed deep sequencing and analysed data. K.S. and E.A. performed single-cell genotyping. J.W., J.Q., C.R.E., W.T.B., J.L., E.N.D., P.G., J.M.C., G.H.L., P.M. and J.C.I.B. supervised the project or related experiments. K.S., Y.T., J.W. and J.C.I.B. wrote the manuscript with input from all the authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C.I.B. (belmonte@salk.edu).

Reviewer Information *Nature* thanks M. Porteus and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Plasmids. Vector expressing both gRNA and mCherry (pCAGmCherry-gRNA) was generated as previously described³⁰. To construct gRNA expression vectors, each 20 bp target sequence was sub-cloned into pCAGmCherry-gRNA or gRNA Cloning Vector (Addgene 41824). The CRISPR/Cas9 target sequences (20 bp target and 3 bp PAM sequence (underlined)) used in this study include: scramble, GCTTAGTTACGCGTGGACGAAGG; mutant GFP, CAGG GTAATCTCGAGAGCTTAGG; MH1, GCCGCTTACTTAGGTCCCCGGG; and MH2, GGAGATCCACTCTCGAGCCCCGGG; for PITCh donor: mouse *Tubb3*, AGCTGGAGCAACTTCACTTGGG; human *TUBB3*, AGCTGGAGCAGCTTCACTTGGG; human *KCNQ1*, AGTACGTGGCCTCTGGGGGGCGG; the downstream of CAG promoter in Ai14 mouse, TAGGAAGTCTTAGGGCCCCGGG; rat *Mertk* for HITI, GAGGACCACTGCAACGGGGCTGG; rat *Mertk* for HDR, TCAGGTGCTTAGGCATTTCTGG. The Scramble-gRNA target sequence we designed is an artificial sequence that does not exist in human, mouse and rat genomes. We used the off-target finder software Cas-OFFinder (<http://www.rgenome.net/cas-offinder/>) to confirm that there were no genomic target sites within 2-bp mismatches. We have confirmed that the Scramble-gRNA can cut its target site in the donor vector (Extended Data Fig. 1b). pMDLg/pRRE, pRSV-Rev and pMD2.G (Addgene 12251, 12253 and 12259) were used for packaging lentiviruses. pEGIP*35 and tGFP (Addgene 26776 and 26864) were used for examining HDR and HITI efficiencies. To construct IRESmCherry-0c, IRESmCherry-1c, IRESmCherry-2c, IRESmCherry-MH, IRESmCherry-HDR-0c and IRESmCherry-HDR-2c, IRES and mCherry sequences were amplified with Cas9 target sequence by PCR from pEGIP*35 and pCAGmCherry-gRNA, respectively and co-integrated into pCR-bluntII vector (Invitrogen) with zero, one or two CAS9/gRNA target sequences. Cas9 expression plasmid (hCas9) was purchased from Addgene (41815). To generate different NLS-dCas9 constructs, pMSCV-LTR-dCas9-VP64-BFP (Addgene 46912) was used to amplify dCas9, which was subsequently subcloned into pCAG-containing plasmid with different NLS and 3×Flag tag. To construct pCAG-Cas9 (no NLS), pCAG-1NLS-Cas9-1NLS and pCAG-1BP-NLS-Cas9-1BP-NLS, D10A and H840A mutations of dCas9 plasmids were exchanged to wild-type sequence by In-Fusion HD Cloning kit (Clontech). Then, pCAG-Cas9-2AGFP (no NLS), pCAG-1NLS-Cas9-1NLS-2AGFP and pCAG-1BP-NLS-Cas9-1BP-NLS-2AGFP were constructed by adding 2AGFP downstream of Cas9. To construct pCAG-floxSTOP-1BP-NLS-Cas9-1BP-NLS, 1BP-NLS-Cas9-1BP-NLS was amplified by PCR and exchanged with GFP of pCAG-floxSTOP-EGFP-N1 vector³¹. To construct HITI donor plasmids for mouse and human *Tubb3* gene (*Tubb3*-1c, *Tubb3*-2c, *Tubb3*-2cd, hTUBB3-1c and hTUBB3-2c) and PITCh donor (*Tubb3*-MH), GFP was subcloned into pCAG-floxSTOP plasmid with one or two CAS9/gRNA target sequences. To construct HDR donor for mouse *Tubb3* gene (*Tubb3*-HR), GFP, 5' and 3' homology arms were amplified from pCAG-GFP-N1 or mouse genome, then subcloned into pCAG-floxSTOP plasmid. pCAG-ERT2-Cre-ERT2 was purchased from Addgene (13777). PX551 and PX552 were purchased from Addgene (60957 and 60958). To construct AAV-Cas9, nEF (hybrid EF1α/HTLV) promoter (Invivogen) was exchanged with Mecp2 promoter of PX551. To construct donor/gRNA AAVs for HITI, donor DNA sandwiched by Cas9/gRNA target sequence, gRNA expression cassette and GFPKASH (or mCherryKASH) expression cassettes were subcloned between ITRs of PX552, and generated pAAV-mTubb3, pAAV-Ai14-HITI, pAAV-Ai14-luc, pAAV-Ai14-scramble and pAAV-rMertk-HITI. For pAAV-rMertk-HITI, exon 2 of rat *Mertk* gene including the surrounding intron is sandwiched by Cas9/gRNA target sequence, which is expected to integrate within intron 1 of *Mertk* by HITI. For HDR AAV (pAAV-Ai14-HDR and pAAV-rMertk-HDR), the homology arms were amplified by PCR from mouse and rat genome DNA, and subcloned into AAV backbone plasmid. The plasmids described in this manuscript will be available to academic researchers through Addgene.

Genomic DNA extraction and genomic PCR. Genomic DNAs were extracted using Blood & Tissue kit (QIAGEN) or PicoPure DNA Extraction Kit (Thermo Fisher Scientific). Genomic PCRs were performed using PrimeSTAR GXL DNA polymerase (Takara).

Bisulphite sequencing. Genomic DNA from the transfected HEK293 lines was extracted and bisulphite converted using the Zymo EZ DNA methylation-direct Kit (Zymo Research). The DNA methylation profile of mCherry was analysed by TOPO cloning as described previously³².

Cell lines. H1 hES cells were purchased from WiCell Research, and maintained in hES cell medium³³. HEK293 cell was purchased from ATCC. Cell lines were authenticated by STR analysis. Mycoplasma contamination was checked every 2 months and was found to be negative in all cell lines used.

AAV production. All AAVs were packaged with serotypes 8 or 9 and were generated by the Gene Transfer Targeting and Therapeutics Core (GT3) at the Salk Institute for biological studies.

Animals. ICR, C57BL/6 and ROSA-LSL-tdTomato (known as Ai14)²⁰ mice were purchased from the Jackson Laboratory. Some timed pregnant ICR mice were purchased from SLC Japan (Sizuoka, Japan). RCS and Brown Norway rats were purchased from the Jackson Laboratory. All mice used in this study were from mixed gender, mixed strains and P1 to 12 weeks old. All mouse experiments were approved by the IACUC committee or the RIKEN Center for Developmental Biology and conform to regulatory standards. All rat procedures were conducted with the approval and under the supervision of the Institutional Animal Care Committee at the University of California San Diego and adhered to the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research. The midday of the vaginal plug was designated as embryonic day 0.5 (E0.5).

Minicircle DNA vectors. Construction and production of minicircle DNA vectors were performed as previously described³⁴. Briefly, to construct pre-minicircle plasmids (pIRESmCherry-MC, pIRESmCherry-MC-scramble, pTubb3-MC, pTubb3-MC-scramble, pAi14-GFPNLS-MC, pAi14-GFPNLS-MC-scramble, pAi14-luc-MC and pAi14-luc-MC-scramble), IRESmCherry, GFP or luciferase genes with Cas9/gRNA targeting sequence were cloned into ApaI and SmaI sites of the minicircle producer plasmid pMC.BESPX (a gift from M. Kay, Stanford University School of Medicine). The final minicircle constructs were introduced into the *E. coli* strain 3S2T (a gift from M. Kay) and amplified overnight in Terrific Broth (pH 7.0) (Fisher Scientific). The minicircle production was induced by mixing the overnight TB culture with an equal volume of minicircle induction mix comprising fresh LB and 20% L-arabinose (SBI), followed by a 5 h incubation at 32 °C with shaking at 250 r.p.m. Minicircle DNA was isolated with EndoFree Plasmid Mega Kit (QIAGEN) following the manufacturer's protocol except that the volumes of P1, P2 and P3 buffers were doubled.

Surveyor assay. To confirm the function of the Scramble-gRNA, we performed Surveyor assay in GFP-correction HEK293 line. Briefly, Cas9, Scramble-gRNA, and different donor DNA (IRESmCherry-MC or IRESmCherry-MC-scramble) were transfected into GFP-correction HEK293 line. Three days later, genomic DNA was extracted with DNeasy Blood & Tissue kit. To examine the activity of the generated nuclear localized Cas9, we performed Surveyor assay in human H1 ES cells. Briefly, each 1.5×10^7 feeder-free cultured H1 ES cells were dissociated by TrypLE (Invitrogen), and resuspended in 1 ml of MEF-conditioned medium containing 10 μ M ROCK inhibitor Y-27632 (Biomol Inc.). Cells were electroporated with 25 μ g of pCAGmCherry-KCNQ1 and 25 μ g of different Cas9 (pCAG-Cas9-2AGFP, pCAG-1NLS-Cas9-1NLS-2AGFP or pCAG-1BP-NLS-Cas9-1BP-NLS-2AGFP), and were plated onto 100-mm dishes pre-coated with Matrigel. Two days after electroporation, the cells were dissociated by TrypLE, and Cas9 and gRNA expressing cells were sorted out as GFP/mCherry double-positive cells by BD influx cell sorter (BD), and genomic DNA extracted with DNeasy Blood & Tissue kit. The extracted genomic DNA from the transfected GFP-correction HEK293 line and human H1 ES cells were used for Surveyor assay with SURVEYOR Mutation Detection Kits (Transgenomic) as described previously³⁵.

Generation of GFP-correction HEK293 line. To assess the knock-in efficiency in dividing cells and optimize the HITI method, we established a mutated GFP gene-based reporter system in HEK293. Briefly, pEGIP*35 was co-transfected with pMDLg/pRRE, pRSV-Rev and pMD2.G, packaged and purified as lentiviral vectors according to a published protocol³⁶. HEK293 cells were transduced in suspension with lentiviral EGIP*35 vector and 4 μ g ml⁻¹ polybrene for 1 h. After brief centrifugation to remove any residual lentiviral vector, the cells were seeded in 100-mm dishes. Three days after transduction, puromycin (1–2 μ g ml⁻¹; Invitrogen) was added to the medium. After 10 days, single colonies were individually picked up and expanded as GFP-correction HEK293 line.

Culture of mouse primary neurons. Primary neurons were obtained from the cortex of E14.5 ICR mouse brains. After the embryo retrieval, all dissection procedures were performed in a cold solution of 1× phosphate-buffered saline (PBS) with 2% glucose (Gibco). Cortical tissue was dissociated by trypsinization, and 1.5×10^5 cells cm⁻² were plated over coated poly-D-lysine coverslips (Neuvitro) with Neurobasal media (Gibco) supplemented with 2% B27 (Gibco) and 0.25% Glutamax (Gibco). The cultures were incubated at standard conditions (37 °C in humidified 5% CO₂/95% air atmosphere). Half volume of culture media was replaced every 3 days.

Differentiation and culture of human ES cell-derived pan neurons. The differentiation protocol from human ES cells to pan neurons was described previously³⁷.

Transfection of *in vitro* cultured cells. Lipofectamine 3000 (Invitrogen), CombiMag Reagent in combination with Lipofectamine 2000 (OZBiosciences)

and DNA-In Neuro Transfection Regent (Amsbio) were used for transfection of HEK293 cells, mouse primary cells and human ES cell-derived pan neurons, respectively. Transfection complexes were prepared following the manufacturer's instructions.

Measurement of targeted gene knock-in efficiency in GFP-correction HEK293 line. To measure the targeted gene knock-in efficiency in GFP-correction HEK293 line, we co-transfected hCas9, gRNA (mutant GFP-gRNA, Scramble-gRNA, MH1-gRNA and/or MH2-gRNA) and donor DNA. Promoterless IRESmCherry plasmids with zero, one, or two CRISPR/Cas9 target sites (IRESmCherry-0c, IRESmCherry-1c and IRESmCherry-2c, respectively) and a minicircle donor (IRESmCherry-MC) were used to measure HITI efficiency. HDR-donors (tGFP and IRESmCherry-HDR-0c) were used to measure HDR efficiency. A PITCh-donor (IRESmCherry-MH) was used to measure PITCh efficiency. IRESmCherry-HDR-2c was used as HDR and HITI dual donors. IRESmCherry-0c and IRESmCherry-MC-scramble were used as genome DNA cut only and donor DNA cut only controls, respectively. The Scramble-gRNA target sequence is an artificial sequence that does not exist in both human and mouse genomes. The Scramble-gRNA was transfected with IRESmCherry-MC-scramble. The MH1-gRNA and MH2-gRNA were co-transfected with IRESmCherry-MH. For other donor shown in Fig. 1a, the mutant GFP-gRNA was co-transfected. The efficiencies of targeted gene knock-in via HDR, PITCh and HITI were determined by calculating the percentage of GFP⁺ or mCherry⁺ cells by FACS LSR Fortessa (BD) and the percentages of PITCh, HITI (without indel) or HITI (with indel) per mCherry⁺ cells were determined by Sanger sequencing.

Isolation of the genome-edited GFP-correction HEK293 clones. The transfected cells were separated into mCherry⁺ and mCherry⁻ populations by FACS via BD Influx (BD), and ~500 cells were plated onto 100-mm dishes pre-coated with wild-type HEK293 cells. Two days after transduction, puromycin (2 µg ml⁻¹; Invitrogen) was added to the medium. After 2 weeks, genome-edited HEK293 clones were manually picked and further analysed by PCR and sequencing to determine the genotype.

Immunocytochemistry of primary neurons. Cells were fixed in 4% paraformaldehyde (PFA) at room temperature for 15 min. Then cells were blocked and permeabilized with 5% Bovine Serum Albumin (BSA) and 0.1% Triton X-100 in PBS for 50 min with shaking at room temperature. Primary antibodies were diluted in 2.5% BSA/PBS and cells were incubated overnight at 4 °C in a wet chamber with anti-GFP (Aves) and anti-βIII tubulin (Sigma) antibodies. Next day, cells were washed with 0.2% Tween 20 in PBS, and incubated for 1 h at room temperature with the secondary antibodies Alexa Fluor 488 (Thermo Fisher) or Alexa Fluor 647 (Thermo Fisher). After a second round of washing with 0.2% Tween 20 in PBS, the cells were mounted using DAPI-Vector Shield mounting media (Vector) and stored at 4 °C. To examine cell proliferation status, we added 2 µM EdU (Invitrogen) in the transfected neurons, and detected EdU positive cells by Click-iT EdU kit (Invitrogen).

Immunocytochemistry of primary tissues. Animals were harvested after transcardial perfusion using PBS followed by 4% PFA. Organs were dissected out and post-fixed with 2% PFA and 15% sucrose in PBS at 4 °C for 16–20 h, then immersed in 30% sucrose in PBS at 4 °C before sectioning. Mouse brains were fixed in 1% PFA in 0.1 M phosphate buffer (pH 7.4) at 4 °C for 24 h followed by cryoprotection in 25% sucrose overnight at 4 °C. For neonatal brain, brains were embedded in OCT compound (Sakura Tissue-Tek) and sectioned by Cryostat (14 µm). Well-dried sections were washed 3 times with PBST (1% Tween 20 in PBS) and treated with blocking buffer (2% donkey serum and 0.2% Triton X-100 in PBS, pH 7.4) for 1 h at room temperature, followed by incubation with primary antibodies diluted in the same buffer overnight at 4 °C. The primary antibodies used were Anti-GFP (Aves) and anti-mCherry (Abcam). Sections were washed three times in PBST and treated with secondary antibodies conjugated to Alexa Fluor 488 or Alexa Fluor 546 (Thermo Fisher) for 1 h at room temperature. After wash, the sections were mounted with mounting medium (PermaFluor, Thermo scientific). For adult brain, 50 µm coronal brain sections were prepared using a freezing microtome and stored in PBS with 0.01% sodium azide at 4 °C. Free-floating sections were incubated at 4 °C for 16–48 h with goat anti-GFP (Rockland) primary antibodies in PBS/0.5% normal donkey serum/0.1% Triton X-100, followed by the appropriate secondary antibodies conjugated with Alexa Fluor 488 at room temperature for 2–3 h. Sections were counterstained with 10 µM DAPI in PBS for 30 min to visualize cell nuclei. Immunostained tissue sections were mounted on slides with polyvinyl alcohol mounting medium containing DABCO and allowed to air-dry overnight. For other tissues, the harvested tissues were embedded in OCT compounds and frozen. Serial or axial frozen sections (thickness 10–20 µm) were prepared using a cryostat, which were then placed on silanized slides and air-dried. The sections were washed with PBS, followed by 1 h room temperature incubation by blocking buffer containing 3% normal goat serum, 0.3% and Triton X-100 in PBS, then incubated with the first antibody solution overnight. The primary antibodies used

were anti-GFP, anti-mCherry, anti-dystrophin (Sigma), anti-actin, anti-smooth-muscle antibody (Sigma) and anti-human-serum-albumin antibody (R&D). After wash, the sections were immunostained with secondary antibody solution for 1 h at room temperature. The secondary antibodies used were Alexa Fluor 488, 568 or 647. After sequential washing with 0.2% Tween 20/PBS, 0.05% Tween 20/PBS, and PBS, the sections were mounted with DAPI Fluoromount-G (Southern Biotech). For rat, retinal cryosections were rinsed in PBS and blocked in 0.5% Triton X-100 in 5% BSA in PBS for 1 h at room temperature. Anti-Merck antibody (eBioscience) was diluted in 5% BSA in PBS and incubated with sections overnight at 4 °C. The sections were then washed three times with PBS, incubated with IgG secondary antibody tagged with Alexa Fluor 555 (Thermo Fisher) in PBS at room temperature for 1 h, and washed with PBS. Cell nuclei were counterstained with DAPI. Sections were mounted with Fluoromount-G (Southern Biotech) and coverslipped. Images were captured by Keyence BZ-9000 microscope.

Nuclear/cytoplasm ratio. To measure intracellular localization of dCas9, we followed a previous report¹⁶. In brief, the dCas9-transfected HEK293 cells were fixed with 4% PFA and stained with anti-Flag (Sigma) and DAPI (Vector). The intensity of fluorescence was measured using the PlotProfile tool of ImageJ software. Values were obtained independently in cytoplasmic and nuclear compartments in single transfected cells. Relative fluorescence values of nuclear intensity were divided by the values found in cytoplasm to obtain the nuclear/cytoplasm ratio.

Gene transfer into mouse embryos by *in utero* electroporation. The experimental procedures for *in utero* electroporation have been described previously³⁸. E15.5 pregnant ICR mice were anaesthetized by 500 µl IP injection of 10% Nembutal (Dainippon sumitomo kagaku). 1 µl of DNA mixture, containing the pCAG-1BPnLS-Cas9-1BPnLS (0.5 µg µl⁻¹), mouse *Tubb3* gene target pCAGmCherry-gRNA (0.5 µg µl⁻¹) and either donor cut-only control donor (Tubb3-MC-scramble), minicircle donor (Tubb3-MC), 2-cut (Tubb3-2c) or HDR donor (Tubb3-HDR) vectors (0.8 µg µl⁻¹) was injected into the hemisphere of the fetal brain. For visually confirming the injection, 0.005% fast green solution (Wako) was mixed with the DNA. Fetuses were tweezed by paddles of the tweezer electrodes (CUY21 electroporator, NEPA GENE). For tamoxifen (TAM) inducible Cre-dependent Cas9 expression system, fetuses were injected with 1 µl of DNA mixture into the hemisphere, containing the pCAG-floxSTOP-1BPnLS-Cas9-1BPnLS (0.5 µg µl⁻¹), pCAG-ERT2CreERT2 (0.5 µg µl⁻¹), pCAG-mcherry-U6-gRNA (0.5 µg µl⁻¹) and either minicircle donor (Tubb3-MC) or HDR donor (Tubb3-HDR) vectors (0.8 µg µl⁻¹). 50 µl of 10 mg ml⁻¹ tamoxifen (Sigma) dissolved in corn oil were injected to P10 and P11 electroporated pups for induction of the Cas9 expression. The GFP knock-in efficiency was measured by the percentage of GFP⁺ cells among transfected cells (mCherry⁺).

RT-PCR from neonatal mouse brain. The DNA mixtures were transfected by *in utero* electroporation at E15.5 of mouse brain and the mice were euthanized at P10. The collected brains from P10 mice were trypsinized for 40 min at 37 °C, then dissociated to single cells by pipetting. About 22,000 electroporated cells were collected by FACS sorting (SH-800, Sony). Total RNA was extracted from the sorted cells with RNeasy mini kit (Qiagen) and cDNA was synthesized by SuperScript VIL0 (Invitrogen). RT-PCR was performed with PrimeSTAR GXL polymerase as following the manufacturer's protocol with 10% of 5 M betaine solution (Sigma).

***In vivo* muscle electroporation.** The DNA mixture for Scramble control (25 µg of pCAG-1BPnLS-Cas9-1BPnLS, 25 µg of Scramble-gRNA-mCherry and 10 µg of Ai14-GFPnLS-MC-scramble), without Cas9 (25 µg of empty vector, 25 µg of Ai14gRNA-mCherry and 10 µg of Ai14-luc-MC or Ai14-GFP-MC) and with Cas9 (25 µg of pCAG-1BPnLS-Cas9-1BPnLS, 25 µg of Ai14gRNA-mCherry and 10 µg of Ai14-luc-MC or Ai14-GFP-MC) were prepared in 25 µl TE. Wild-type or Ai14 mice were anaesthetized with intraperitoneal injection of ketamine (100 mg kg⁻¹) and xylazine (16 mg kg⁻¹). For quadriceps muscle electroporation, a small portion of the quadriceps muscle was surgically exposed in the hind limb. Plasmid DNA mixture was injected into the muscle using a 29-gauge insulin syringe. One minute following plasmid DNA injection, a pair of electrodes was inserted into the muscle to a depth of 5 mm to encompass the DNA injection site and muscle was electroporated using an Electro Square Porator T820 (BTX Harvard Apparatus). Electrical stimulation was delivered twenty pulses at 100 V for 20 ms. After electroporation, skin was closed and mice were recovered on a 37 °C warm pad. For panniculus carnosus muscle electroporation, the hair of back skin was depilated with depilatory cream. The above mixture of DNA solutions were conjugated and subcutaneously injected to right and left side, respectively. The injected areas of skin and subcutaneous tissue were vertically sandwiched by plate-and-fork type electrodes, consist of a pair of stainless-steel tweezers, one with a rectangular plate, 10 mm long and 5 mm wide, and the other with a fork consisting of three straight needles at 2.5 mm intervals, which are 10 mm long and 0.5 mm in diameter. The interface of skin and the rectangular electrode was covered with electroconductive gel (SpectraGel 360, Parker Labs). Twenty 18 V/50 ms/1 Hz square pulses followed

by another 20 pulses of the opposite polarity were delivered using Electro Square Porator T820. Two weeks after the electroporation, mice were euthanized, and tissues were obtained.

Tissue pressure-mediated transfection. The DNA mixture without Cas9 (100 µg of empty vector, 100 µg of Ai14gRNA-mCherry and 50 µg of Ai14-luc-MC) and with Cas9 (100 µg of pCAG-1BPNS-Cas9-1BPNS, 100 µg of Ai14gRNA-mCherry and 50 µg of Ai14-luc-MC) were prepared in 200 µl saline. A midline laparotomy was performed and the right kidney of wild-type or Ai14 mouse was exteriorized. After exposure of kidney, mice were intravenously injected with plasmid DNA mixture, immediately followed by pressing the right kidney gripped between thumb and index finger 20 times for a period of 1 s each as described previously³⁹.

In vivo electroporation for kidney. The DNA mixture for Scramble control (100 µg of pCAG-1BPNS-Cas9-1BPNS, 100 µg of Scramble-gRNA-mCherry and 50 µg of Ai14-GFPNLS-MC-scramble), without Cas9 (100 µg of empty vector, 100 µg of Ai14gRNA-mCherry and 50 µg of Ai14-GFP-MC) and with Cas9 (100 µg of pCAG-1BPNS-Cas9-1BPNS, 100 µg of Ai14gRNA-mCherry and 50 µg of Ai14-GFP-MC) were prepared in 200 µl saline. A midline laparotomy was performed. The right kidney of Ai14 mouse was exteriorized and subsequently decapsulated, leaving the adrenal gland intact. The exposed kidney was pricked with electrode needles after injection of plasmid DNA mixture from tail vein and subsequently received electroporation 100 V, 50 ms pulse, six times using an Electro Square Porator T820.

Luciferase detection. Mice were examined at 2 weeks after DNA transfection or electroporation by BLI performed using an IVIS Kinetic 2200 (Caliper Life sciences). Mice were IP injected with 150 mg kg⁻¹ D-Luciferin (BIOSYNTH), anaesthetized with isoflurane and dorsal images were then captured 10 min post luciferin injection.

AAV infection in mouse primary neuron. Primary cultures of neurons were used after three days in culture, the AAV solution (without Cas9, AAV-mTub3 (1.5 × 10¹⁰ GC); with Cas9, AAV-Cas9 (1.5 × 10¹⁰ GC) and AAV-mTub3 (1.5 × 10¹⁰ GC)) was added and cultures were kept at standard conditions for 5 days, following immunocytochemistry or DNA extraction.

Stereotax AAV injection in adult brain. C57BL/6 mice received AAV8 injections at P75. We used 1:1 mixture of AAV-Cas9 (1.5 × 10¹³ GC ml⁻¹) and AAV-mTub3 (2.3 × 10¹³ GC ml⁻¹). As a control, 1:1 mixture of AAV-mTub3 and HBSS buffer was used. Mice were anaesthetized with 100 mg kg⁻¹ of ketamine and 10 mg kg⁻¹ of xylazine cocktail via intra-peritoneal injections and mounted in a stereotax (David Kopf Instruments Model 940 series) for surgery and stereotaxic injections. Virus was injected into the centre of V1, using the following coordinates: 3.4 mm rostral, 2.6 mm lateral relative to bregma and 0.5–0.7 mm ventral from the pia. We injected 200 nl of AAVs using air pressure by picospritzer (General Valve Corp). To prevent virus backflow, the pipette was left in the brain for 5–10 min after completion of injection. Mice were housed for two weeks to allow for gene knock-in.

Intramuscular AAV injection. Ai14 mice were anaesthetized with intraperitoneal injection of ketamine (100 mg kg⁻¹) and xylazine (16 mg kg⁻¹). A small portion of the quadriceps muscle was surgically exposed in the hind limb. The AAV8 mixture (without Cas9, AAV-Ai14-HITI (1.5 × 10¹⁰ GC); with Cas9, AAV-Cas9 (1.5 × 10¹⁰ GC) and AAV-Ai14-HITI (1.5 × 10¹⁰ GC)) was injected into the quadriceps muscle using a 29 Gauge insulin syringe. After AAV injection, skin was closed and mice were recovered on a 37 °C warm pad.

Intravenous (IV) AAV injection. The newborn (P1) of Ai14 mice were used for IV AAV8 or AAV9 injection following a previous report⁴⁰. The AAV mixtures (without Cas9, AAV-Ai14-HITI (5 × 10¹⁰ or 2 × 10¹¹ GC); with Cas9, AAV-Cas9 (5 × 10¹⁰ or 2 × 10¹¹ GC) and AAV-Ai14-HITI (5 × 10¹⁰ or 2 × 10¹¹ GC); Scramble control, AAV-Cas9 (5 × 10¹⁰ GC) and AAV-Ai14-Scramble (5 × 10¹⁰ GC); HDR, AAV-Cas9 (5 × 10¹⁰ GC) and AAV-Ai14-HDR (5 × 10¹⁰ GC)) were injected via temporal vein of the P1 mouse.

Tail vein AAV injection. The AAV8 mixtures (without Cas9, AAV-Ai14-luc (2 × 10¹¹ GC); with Cas9, AAV-Cas9 (2 × 10¹¹ GC) and AAV-Ai14-luc (2 × 10¹¹ GC)) were injected via tail vein for luciferase knock-in. The AAV mixtures (without Cas9, AAV-Ai14-HITI (2 × 10¹¹ GC); with Cas9, AAV-Cas9 (2 × 10¹¹ GC) and AAV-Ai14-HITI (2 × 10¹¹ GC)) were injected via tail vein for GFP knock-in.

Image capture and measurement of in vivo GFP knock-in efficiency. For immunocytochemical analyses, the cells and tissues were visualized by confocal microscopy using a Zeiss LSM 780 Laser Scanning Confocal or Olympus FV1000 confocal microscope (Olympus). At least five pictures were obtained from each group and animal. We analysed at least three animals. Pictures were analysed with ZEN 2 (blue edition) and NIH ImageJ (FIJI) software. For the mouse primary neurons and human pan neurons analyses, the total number of positive cells for each marker were directly counted with the multi-point tool of NIH ImageJ software. The percentage of GFP⁺ cells was calculated among transfected cells (mCherry⁺) or total cells (DAPI⁺). The intracellular distribution of GFP was observed in around

100 independent events for each condition, where the focused cell was observed at different stacks to determine the presence or absence of GFP at the nucleus space. To assess the efficiency of GFP knock-in in brain after local AAV injection, we counted number of GFP⁺, mCherry⁺ and DAPI⁺ cells of 300 µm within injection sites and determined the GFP knock-in efficiency per infected cells or per cell. To assess the efficiency of GFP knock-in in liver, heart and muscle after systemic AAV injection, we counted the number of GFP⁺ and DAPI⁺ cells and determined the GFP knock-in efficiency per cell.

Single-cell genotyping. To collect GFP-positive single cells from muscle and heart of AAV-injected mouse, animals were harvested after transcardial perfusion using PBS. Organs were dissected out and isolated as single cells with published methods^{41,42}. The single nuclei per cell was confirmed by fluorescent microscope with DAPI staining and separated by BD influx cell sorter. Each single GFP⁺ cell was sorted into 5 µl of lysis buffer from PicoPure DNA Extraction Kit by BD influx cell sorter and used as a template for first round of PCR to amplify the target genome with PrimeSTAR GXL polymerase following the manufacturer's protocol. The first round of PCR product was purified using Agencourt AMPure XP (Beckman Coulter), then subject to second round PCR and sequencing to confirm the genotype. Based on sequencing result of 5' junction end, single cell genotyping can separate biallelic GFP knock-in, monoallelic GFP knock-in with indels at another target, monoallelic GFP knock-in without indels at another target, and unknowns.

Targeted deep sequencing. The top 12 predicted off-target sites were searched using The CRISPR Design Tool⁴³. The on-target and potential off-target regions were amplified using PrimeSTAR GXL DNA polymerase from the liver DNA via IV injection and used for library construction. Equal amounts of the genomic DNA was used to amplify genomic regions flanking the on-target and top 12 predicted off-target nuclease binding sites for library construction. Next, PCR amplicons from previous step were purified using Agencourt AMPure XP, then subject to second round PCR to attach Illumina P5 adapters and sample-specific barcodes. The purified PCR products were pooled at equal ratio for single-end sequencing using Illumina MiSeq at the Zhang laboratory (UCSD). The raw reads were mapped to mouse reference genome mm9 or custom built Ai14 mouse genome using BWA⁴⁴. High quality reads (score >30) were analysed for indel events and Maximum Likelihood Estimate (MLE) calculation as previously described⁴³. As next generation sequencing analysis of indels cannot detect large size deletion and insertion events, CRISPR/Cas9 targeting efficiency and activity shown above is underestimated.

Subretinal injection in rats. Congenic RCS rats at 21 days old were used in the study and divided into three groups. RCS group is non-injection control. The Cas9 + HITI group received a subretinal injection of 2 µl of AAV8 mixture (AAV-Cas9 (1.5 × 10¹⁰ GC) and AAV-rMertk-HITI (1.5 × 10¹⁰ GC)) in the eyes. The Cas9 + HDR group received a subretinal injection of 2 µl of AAV8 mixture (AAV-Cas9 (1.5 × 10¹⁰ GC) and AAV-rMertk-HDR (1.5 × 10¹⁰ GC)) in the eyes. Wild-type rats without an injection served as a normal control. Experimental rats were anaesthetized with an intraperitoneal injection of a mixture of ketamine and xylazine. Pupils were dilated with 1% topical tropicamide. Subretinal injection was performed under direct visualization using a dissecting microscope with a pump microinjection apparatus (Picospritzer III; Parker Hannifin Corporation) and a glass micropipette (internal diameter ~50–75 µm). Two microlitres of AAV mixture was injected into the subretinal space through a small scleral incision. A successful injection was judged by creation of a small subretinal fluid bleb. Fundus examination was performed immediately following injection, and rats showing any sign of retinal damage such as bleeding were discarded and excluded from the final animal counts.

ERG recording. To monitor the efficacy of gene knock-in in vision rescue, ERG studies were performed at 4 weeks after treatment before the animals were euthanized for histology. The dark-adapted ERG response was recorded as described previously⁴⁵. In brief, rats were dark-adapted for 14 h before the commencement of each ERG recording session. They were deeply anaesthetized as described for the surgical procedure above. Eyes were treated with 1% topical tropicamide to facilitate pupillary dilation. Each rat was tested in a fixed state and manoeuvred into position for examination within a Ganzfeld bowl (Diagnosys LLC). One active lens electrode was placed on each cornea, with a subcutaneously placed ground needle electrode positioned in the tail and the reference electrodes placed subcutaneously in the head region approximately between the two eyes. Light stimulations were delivered with a xenon lamp at 0.01 and 0.3 cds m⁻² in a Ganzfeld bowl. For the flicker ERG measurement, rats were adapted at a background light of 10 cds m⁻², and light stimulation was set at 30 cds m⁻². The recordings were processed using software supplied by Diagnosys.

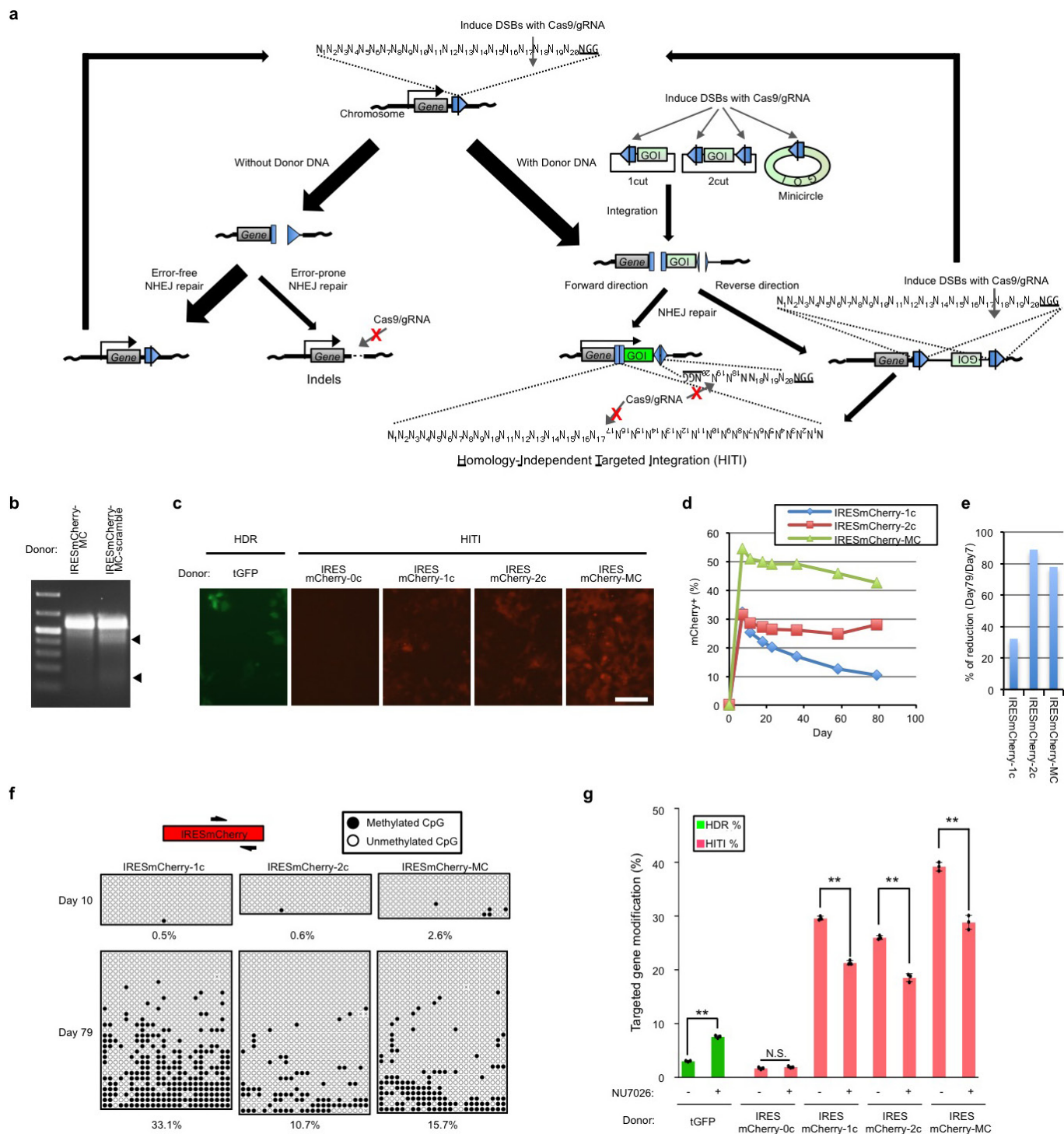
Histological analysis of the rat eye. Following ERG recordings, rats were euthanized and retinal cross-sections were prepared for histological evaluation of ONL preservation. Rats were euthanized with CO₂, and eyeballs were dissected out and

fixed in 4% PFA. Cornea, lens, and vitreous were removed from each eye without disturbing the retina. The remaining retina-containing eyecup was infiltrated with 30% sucrose and embedded in OCT compound. Horizontal frozen sections were cut on a cryostat. Retinal cross-sections were prepared for histological evaluation by staining with haematoxylin and eosin (H&E).

DNA and RNA analyses of the rat eye. Following ERG recordings, rats were euthanized. DNA and RNA were isolated from retina-choroid complex using an AllPrep DNA/RNA Mini Kit (Qiagen). DNA was further used for PCR and TOPO sequencing. cDNA was synthesized from RNA using a Superscript III reverse transcriptase kit (Invitrogen) according to the manufacturer's instructions. Quantitative PCR was performed via 40 cycle amplification using following primers (MertK-F1: GCATTTCGTGGTGGAAAGAT, MertK-R1: TGGGATCAGACACAACCTCTC) and Power SYBR Green PCR Master Mix on a 7500 Real-Time PCR System (Applied Biosystems). Measurements were performed in triplicate and normalized to endogenous GAPDH levels. The relative fold change in expression was calculated using the $\Delta\Delta C_t$ method (C_t values <30).

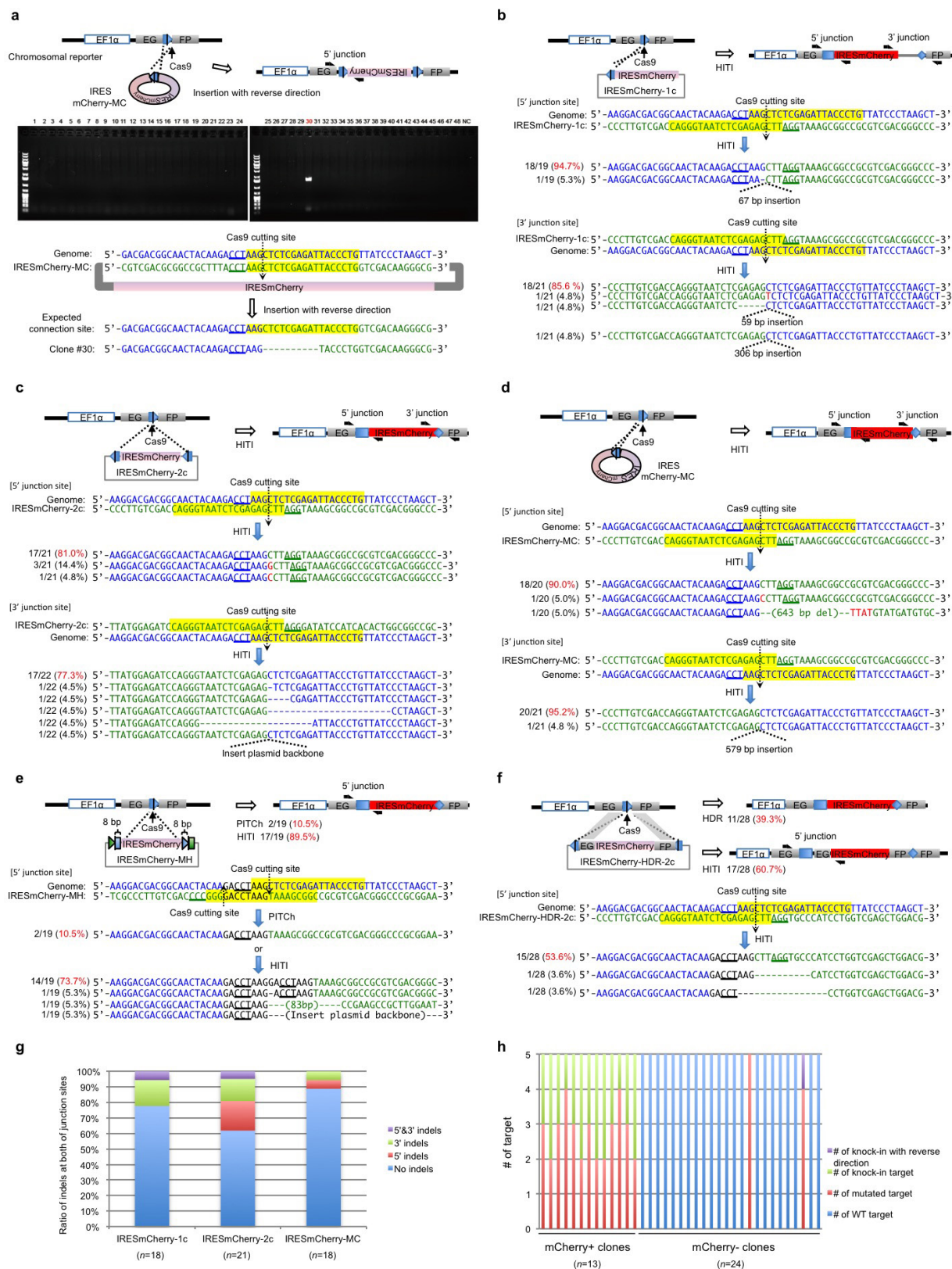
Data availability. The analysis for insertion and deletion (indel) events and Maximum Likelihood Estimate (MLE) calculation were done as previously described⁴³. All custom scripts can be provided upon request. Raw Illumina sequencing reads for this study have been deposited in the National Center for Biotechnology Information Short Read Archive and accessible through SRA accession number SRP069844.

30. Wu, J. *et al.* An alternative pluripotent state confers interspecies chimaeric competency. *Nature* **521**, 316–321 (2015).
31. Shitamukai, A., Konno, D. & Matsuzaki, F. Oblique radial glial divisions in the developing mouse neocortex induce self-renewing progenitors outside the germinal zone that resemble primate outer subventricular zone progenitors. *J. Neurosci.* **31**, 3683–3695 (2011).
32. Liu, G.-H. G. *et al.* Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs. *Cell Stem Cell* **8**, 688–694 (2011).
33. Liu, G.-H. G. *et al.* Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome. *Nature* **472**, 221–225 (2011).
34. Kay, M. A., He, C.-Y. & Chen, Z.-Y. A robust system for production of minicircle DNA vectors. *Nat. Biotechnol.* **28**, 1287–1289 (2010).
35. Sanjana, N. E. *et al.* A transcription activator-like effector toolbox for genome engineering. *Nat. Protocols* **7**, 171–192 (2012).
36. Kutner, R. H., Zhang, X.-Y. & Reiser, J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat. Protocols* **4**, 495–505 (2009).
37. Liu, G.-H. G. *et al.* Modelling Fanconi anemia pathogenesis and therapeutics using integration-free patient-derived iPSCs. *Nat. Commun.* **5**, 4330 (2014).
38. Takahashi, M., Nomura, T. & Osumi, N. Transferring genes into cultured mammalian embryos by electroporation. *Dev. Growth Differ.* **50**, 485–497 (2008).
39. Mukai, H., Kawakami, S. & Hashida, M. Renal press-mediated transfection method for plasmid DNA and siRNA to the kidney. *Biochem. Biophys. Res. Commun.* **372**, 383–387 (2008).
40. Lampe, S. E. G., Kaspar, B. K. & Foust, K. D. Intravenous injections in neonatal mice. *J. Vis. Exp.* e52037 (2014).
41. Pasut, A., Oleynik, P. & Rudnicki, M. A. Isolation of muscle stem cells by fluorescence activated cell sorting cytometry. *Methods Mol. Biol.* **798**, 53–64 (2012).
42. Pinto, A. R., Chandran, A., Rosenthal, N. A. & Godwin, J. W. Isolation and analysis of single cells from the mouse heart. *J. Immunol. Methods* **393**, 74–80 (2013).
43. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
45. Sauvé, Y., Lu, B. & Lund, R. D. The relationship between full field electroretinogram and perimetry-like visual thresholds in RCS rats during photoreceptor degeneration and rescue by cell transplants. *Vision Res.* **44**, 9–18 (2004).



Extended Data Figure 1 | Optimization of donor vectors for HITI in HEK293 cells. **a**, Schematic of NHEJ-mediated targeted genome editing and different HITI donor vectors with *Streptococcus pyogenes* Cas9. Blue pentagon, Cas9/gRNA target sequence. Black line within blue pentagon, Cas9 cleavage site. GOI, gene of interest; DSBs, double strand breaks; Indels, insertions and deletions. **b**, Surveyor nuclease assay performed with Cas9, Scramble-gRNA, and different donor plasmids (IRESmCherry-MC or IRESmCherry-MC-scramble) in the GFP-correction HEK293 line. The two lower bands are cleaved DNA products by Surveyor nuclease, indicating that Cas9/Scramble-gRNA cut scramble target sequence on the donor, but not genomic sequence. **c**, Representative immunofluorescence images of the targeted gene

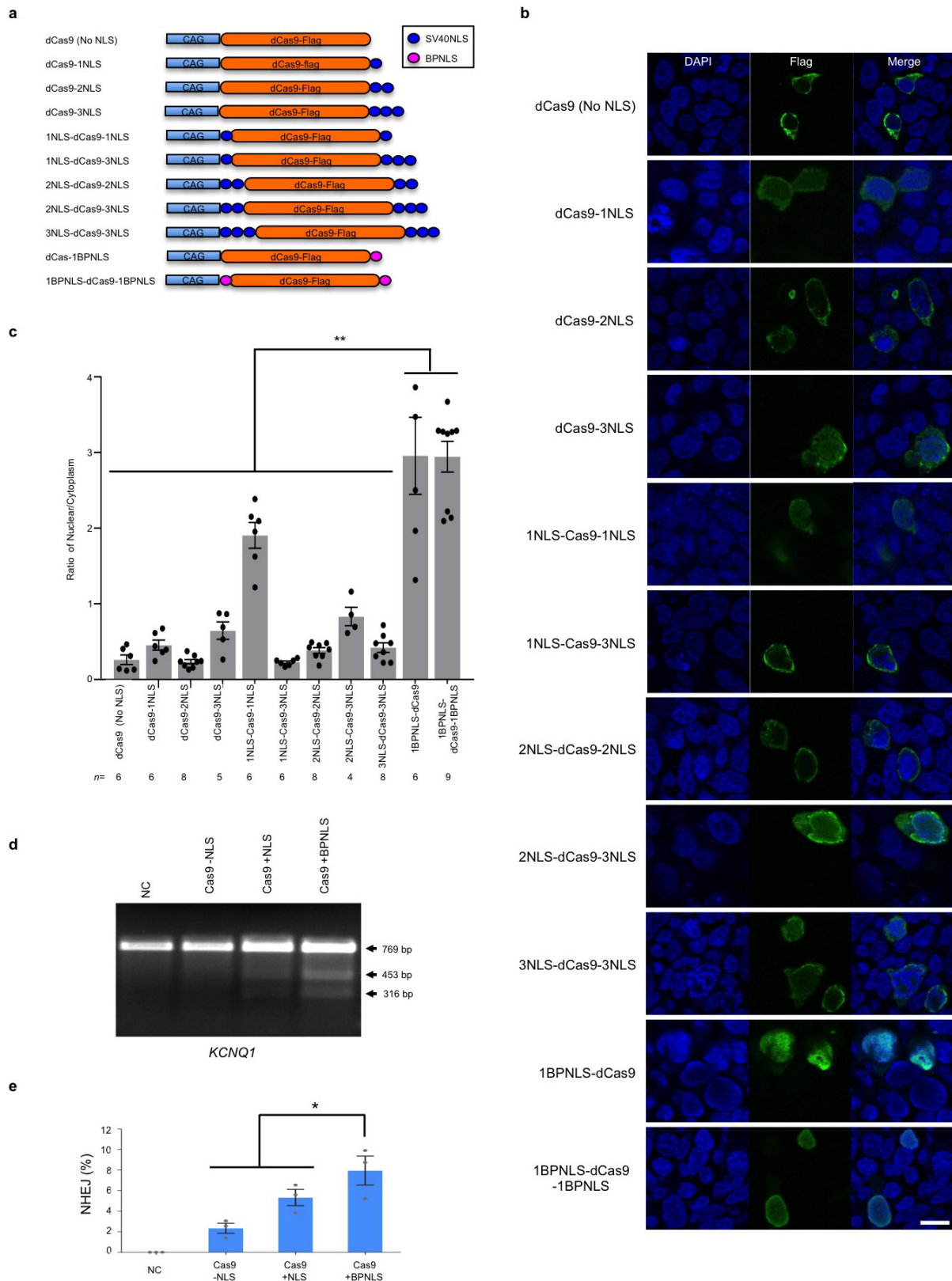
integration in the GFP-correction HEK293 line by HDR and HITI. Scale bar, 100 μ m. **d**, Time course studies of the percentages of mCherry⁺ cells with different HITI donor vectors. **e**, The percentage of mCherry⁺ cells on day 79 versus day 7. **f**, The CpG methylation status of the mCherry gene at early (day 10) and late passages (day 79) with different HITI targeting vectors. Two black half-arrows indicate primers for bisulphite sequencing. **g**, The effect of the NHEJ inhibitor (NU7026; 30 μ M) on knock-in efficiencies by HDR and HITI. Results were obtained from three replicate wells and presented as mean \pm s.d. The input data points were shown as black dots. N.S., not significant. ** $P < 0.01$, unpaired Student's *t*-test. For gel source image, see Supplementary Fig. 1. For source data, see Supplementary Table 5.



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Sequencing analyses of the IRESmCherry knock-in clones via HITI in HEK293 cells. **a**, Analysis of the direction of targeted insertion for HITI with IRESmCherry-MC donor. Detection of reverse integrated IRESmCherry-MC from mCherry⁻ single-cell colonies via PCR. Only one (no. 30, highlighted in red) out of 48 mCherry⁻ clones (non-transfected, non-edited or reverse integrated) was integrated in reverse direction with a 10-bp deletion at junction site revealed by PCR and sequencing analysis. **b**, Sequences of the 5' and 3' junction sites of mCherry⁺ clones after IRESmCherry knock-in by HITI in the GFP-correction HEK293 line with IRESmCherry-1c donor. The blue pentagon and sequence highlighted in yellow indicate the Cas9/gRNA target sequence. The black line within the blue pentagon indicates the Cas9 cleavage site. The PAM sequence is underlined. **c**, Sequences of the 5' and 3' junction sites of mCherry⁺ clones after IRESmCherry knock-in by HITI in the GFP-correction HEK293 line with IRESmCherry-2c donor. **d**, Sequences of the 5' and 3' junction sites of mCherry⁺ clones after IRESmCherry knock-in by HITI in the GFP-correction HEK293 line with

IRESmCherry-MC donor. **e**, Sequences of the 5' junction site of mCherry⁺ clones after IRESmCherry knock-in by PITCh or HITI in the GFP-correction HEK293 line with IRESmCherry-MH donor. **f**, Sequences of the 5' junction site of mCherry⁺ clones after IRESmCherry knock-in by HDR or HITI in the GFP-correction HEK293 line with IRESmCherry-HDR-2c donor. **g**, The fidelity of 5' and 3' junction sites of mCherry⁺ clones after IRESmCherry knock-in by HITI in the GFP-correction HEK293 line with IRESmCherry-1c, IRESmCherry-2c and IRESmCherry-MC donor. **n**, number of analysed clones. **h**, IRESmCherry-MC donor was transfected into the GFP-correction HEK293 line which has five copies of the target sequence in the genome. The ratio of IRESmCherry knock-in and mutation at all target loci of mCherry⁺ clones and the ratio of unmodified, mutated and reverse integrated target loci of mCherry⁻ clones were examined by sequencing. **n**, number of analysed clones. For gel source images, see Supplementary Fig. 1. For source data, see Supplementary Table 6.

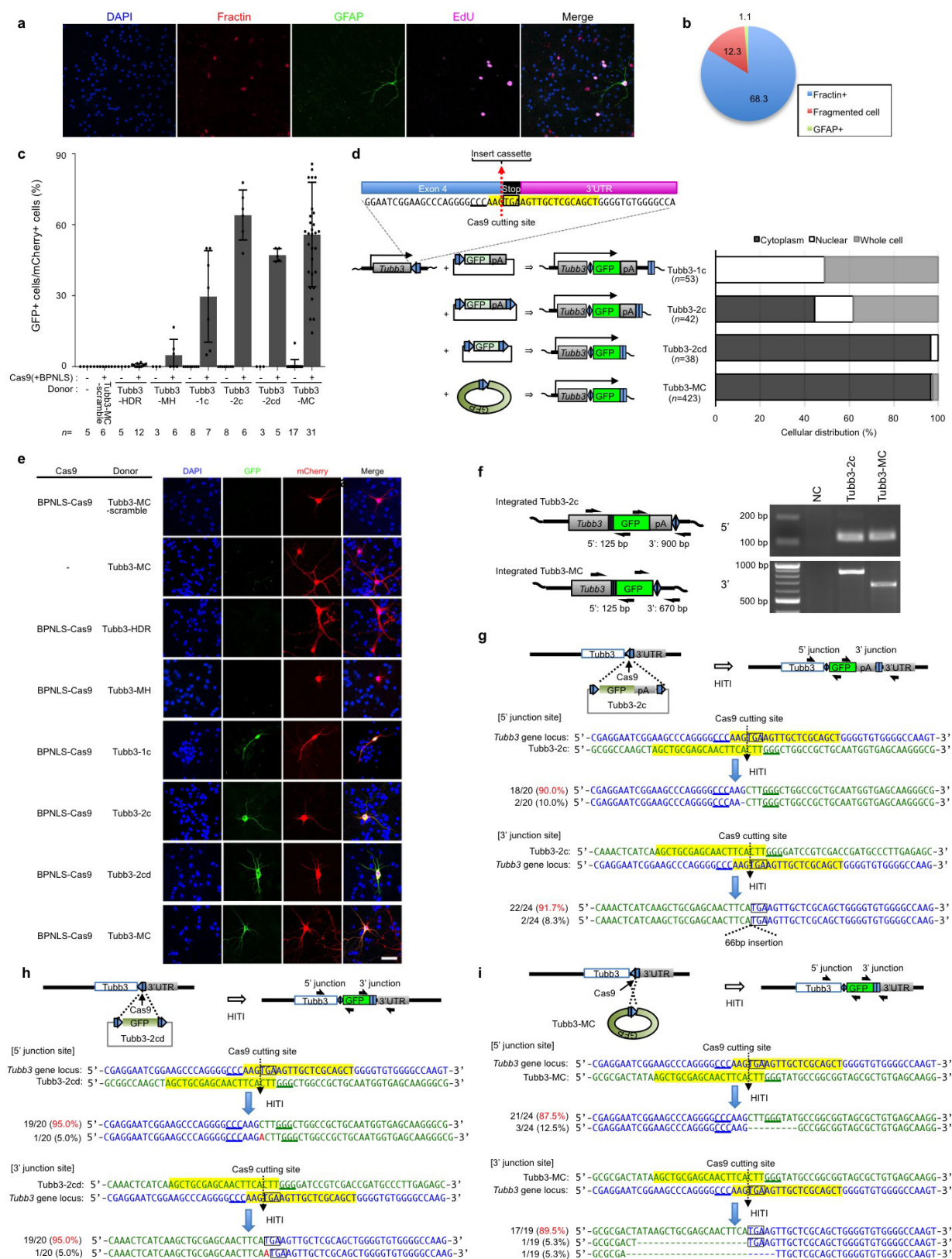


Extended Data Figure 3 | Optimization of nuclear transport of Cas9.

a, Schematic representation of a series of dCas9-Flag constructs with different nuclear localization signals. **b**, Representative immunofluorescence images of the transfected HEK293 cells stained with Flag antibody to visualize dCas9 localization. DNA was counterstained with DAPI. Scale bar, 50 μ m. **c**, The nuclear/cytoplasm ratio of dCas9 with different NLS signals. *n*, number of analysed cells. Results were presented as mean \pm s.e.m. The input data points were shown as black dots. **d**, Characterization of Cas9

nuclease activity in human ES cells. Agarose gel image (**d**) and

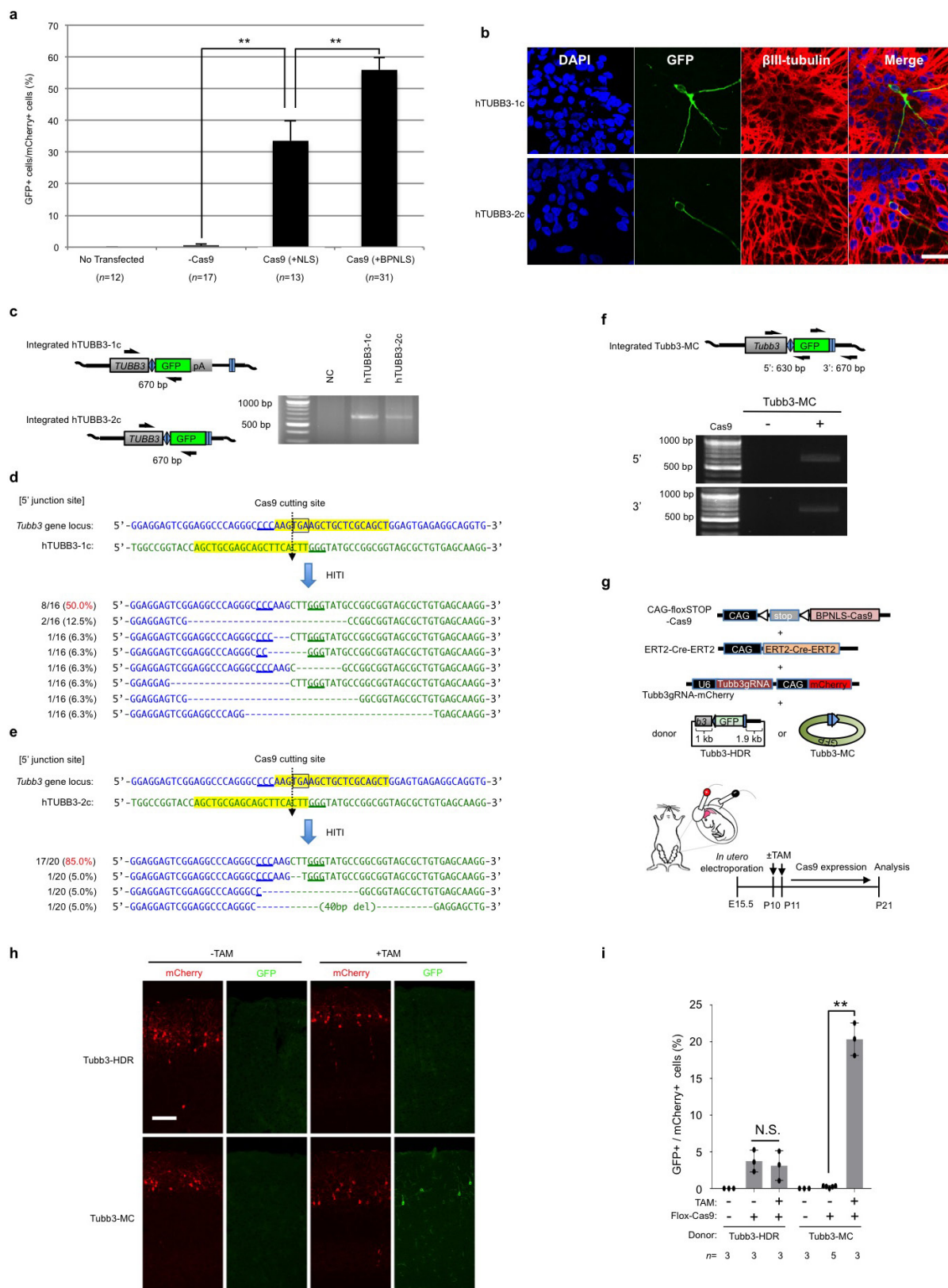
quantification (**e**) show Surveyor nuclease assay performed with the gRNA targeting *KCNQ1* gene and Cas9 with different NLSs, that is, Cas9 -NLS (Cas9 (no NLS)), Cas9 +NLS (1NLS-Cas9-1NLS) and Cas9 +BP-NLS (1BP-NLS-Cas9-1BP-NLS). The two lower bands are cleaved DNA products by Surveyor nuclease. NHEJ (%) indicates the percentage of Cas9/gRNA-mediated gene modification. Results were obtained from three technical replicates and presented as mean \pm s.e.m. The input data points were shown as black dots. **P* < 0.05, paired Student's *t*-test. For gel source image, see Supplementary Fig. 1. For source data, see Supplementary Table 7.



Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | In-depth analyses of GFP knock-in in mouse primary neurons. **a, b**, Characterization of cultured primary neurons. Representative immunofluorescence images (**a**) and quantification (**b**) show that all of the neurons are EdU negative in this culture condition. All EdU⁺ cells were GFAP-positive glia, fractin-positive apoptotic cells, or fragmented cells. Results were obtained from 15 technical replicates and presented as mean. **c**, The percentage of knock-in cells (GFP⁺) per transfected cells (mCherry⁺) with HDR donor (Tubb3-HDR), PITCH donor (Tubb3-MH), 1-cut donor (Tubb3-1c), 2-cut donor (Tubb3-2c), 2-cut donor without polyA (Tubb3-2cd), or minicircle donor (Tubb3-MC). *n*, number of technical replicates. Results were presented as mean \pm s.d. The input data points were shown as black dots. **d**, Left panel, schematic showing inserted DNA sequences, with or without bacteria backbone, with different HITI donors. Blue pentagon and yellow highlight, Cas9/gRNA target sequence. Black line within blue pentagon, Cas9 cleavage site.

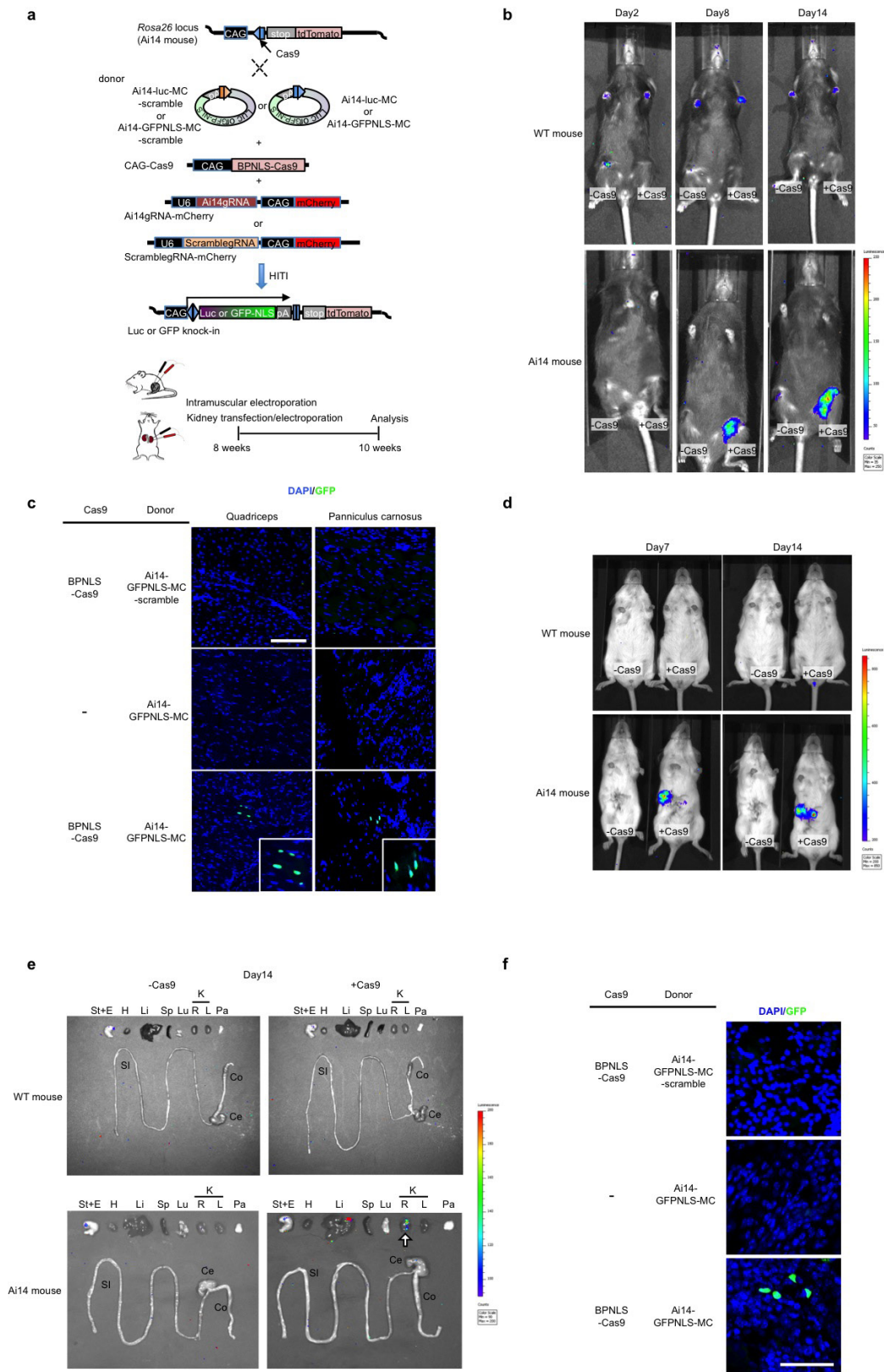
Underlined sequence, PAM sequence. pA, polyA. Right panel, subcellular distribution of TUBB3-GFP. *n*, number of analysed cells. **e**, Representative fluorescence images of the primary neurons transfected with BPNLS-Cas9, gRNA, and different donor plasmids (Tubb3-MC-scramble, Tubb3-HDR, Tubb3-MH, Tubb3-1c, Tubb3-2c, Tubb3-2cd or Tubb3-MC). Different intracellular localization patterns of TUBB3-GFP were observed for different donors. Scale bar, 100 μ m. **f**, Left panel, schematic of knock-in by Tubb3-2c and Tubb3-MC donor at the 5' and 3' ends of *Tubb3* coding region. Black half-arrows indicate PCR primers for detecting integrated sequences. Right panel shows the PCR result. **g–i**, Sequences of the 5' and 3' junction sites after GFP knock-in by HITI in mouse primary neurons with Tubb3-2c (**g**), Tubb3-2cd (**h**) and Tubb3-MC (**i**) donor plasmids. For gel source images, see Supplementary Fig. 1. For source data, see Supplementary Table 8.



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | HITI-mediated GFP knock-in in neurons *in vitro* and *in vivo*. **a**, The percentage of knock-in (GFP⁺) cells in mouse primary neurons per transfected cells (mCherry⁺) with empty vector (–Cas9), Cas9 (+NLS) and Cas9 (+BPNS). *n*, number of technical replicates. Results were presented as mean ± s.e.m. ****P** < 0.01, unpaired Student's *t*-test. **b**, Representative immunofluorescence images of human ES cell-derived pan neurons transfected with BPNS-Cas9, gRNA, and different donor plasmids (hTUBB3-1c or hTUBB3-2c). Scale bar, 100 μm. **c**, PCR analysis of integrated GFP gene at *TUBB3* locus in human ES cell-derived pan neurons. **d**, **e**, Sequences of the 5' junction sites after GFP knock-in by HITI in human ES cell-derived pan neurons with TUBB3-1c donor (**d**) and TUBB3-2c donor (**e**). **f**, Upper panel, schematic of GFP knock-in at the 3' end of the *Tubb3* coding region via Tubb3-MC donor in the neonatal mouse brain. Black half-arrows indicate PCR primers for detecting integrated sequences. Lower panel, genomic PCR results

showing transgene integration at both 5' and 3' ends. **g**, Schematic of *in vivo* targeted GFP knock-in by HITI in the neonatal mouse brain. CAG-floxSTOP-Cas9, inducible BPNS-Cas9 expression plasmid. ERT2-Cre-ERT2, tamoxifen (TAM) inducible Cre expression plasmid. Donor plasmids: Tubb3-HDR or Tubb3-MC. Tamoxifen was injected at P10 and P11. Mice were analysed at P21. **h**, Representative fluorescence images of GFP knock-in at the *Tubb3* locus in neonatal mouse brain by inducible Cas9 expression with HDR donor (Tubb3-HDR) or minicircle HITI donor (Tubb3-MC). Scale bar, 100 μm. **i**, Relative knock-in efficiencies of HDR and HITI donors with or without tamoxifen treatment. *n*, number of pups obtained from two pregnant mice. Results were presented as mean ± s.d. The input data points were shown as black dots. ****P** < 0.01. N.S., not significant. Unpaired student's *t*-test. For gel source images, see Supplementary Fig. 1. For source data, see Supplementary Table 9.



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | HITI via *in vivo* DNA transfection.

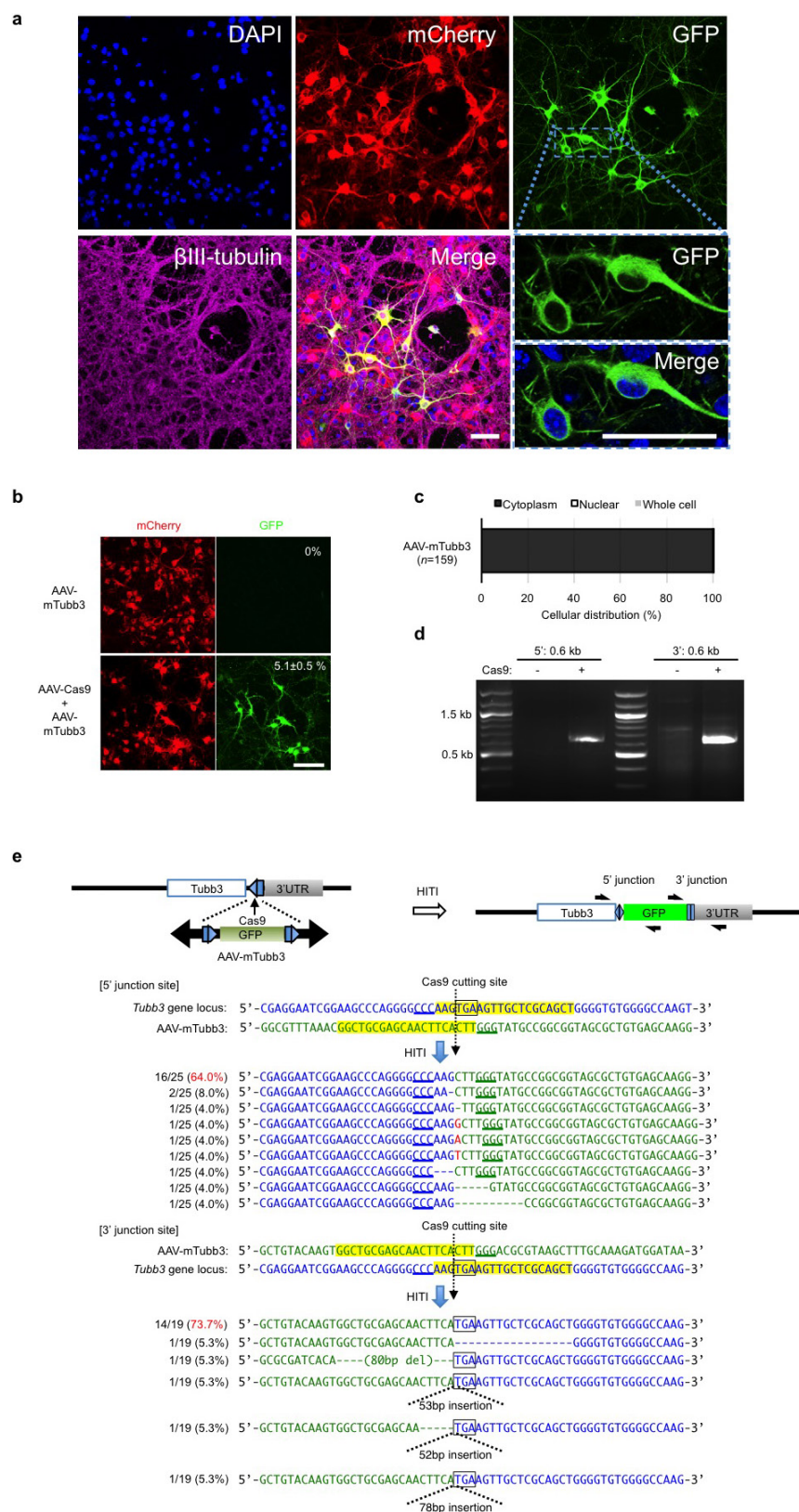
a, Schematic for *in vivo* targeted GFP-NLS or luciferase gene knock-in by HITI. CAG-Cas9, gRNA (Ai14gRNA or Scramble-gRNA)-mCherry and minicircle donor (Ai14-GFPNLS-MC-scramble, Ai14-luc-MC-scramble, Ai14-GFPNLS-MC or Ai14-luc-MC) were locally delivered to mouse kidney or muscle via pressure-mediated transfection and/or electroporation at 8 postnatal weeks and analysed 2 weeks later.

b, *In vivo* imaging of luciferase signals at day 2, day 8, and day 14 post-intramuscular injection of luciferase HITI constructs. Right leg (–Cas9) was injected with empty plasmid, Ai14gRNA-mCherry, and Ai14-luc-MC. Left leg (+Cas9) was injected with CAG-Cas9, Ai14gRNA-mCherry, and Ai14-luc-MC. Top, control wild-type (WT) mouse. Bottom, Ai14 mouse.

c, Representative immunofluorescence images of GFP expression after intramuscular electroporation of GFP-NLS HITI constructs into Ai14 mouse quadriceps (left panel) and panniculus carnosus (right panel). Top panel, donor-cut only control (CAG-Cas9, Scramble-gRNA-mCherry, and Ai14-GFPNLS-MC-scramble). Middle panel, no Cas9 control (empty plasmid, Ai14gRNA-mCherry, and Ai14-GFPNLS-MC). Bottom panel, GFP-NLS HITI (CAG-Cas9, Ai14gRNA-mCherry, and

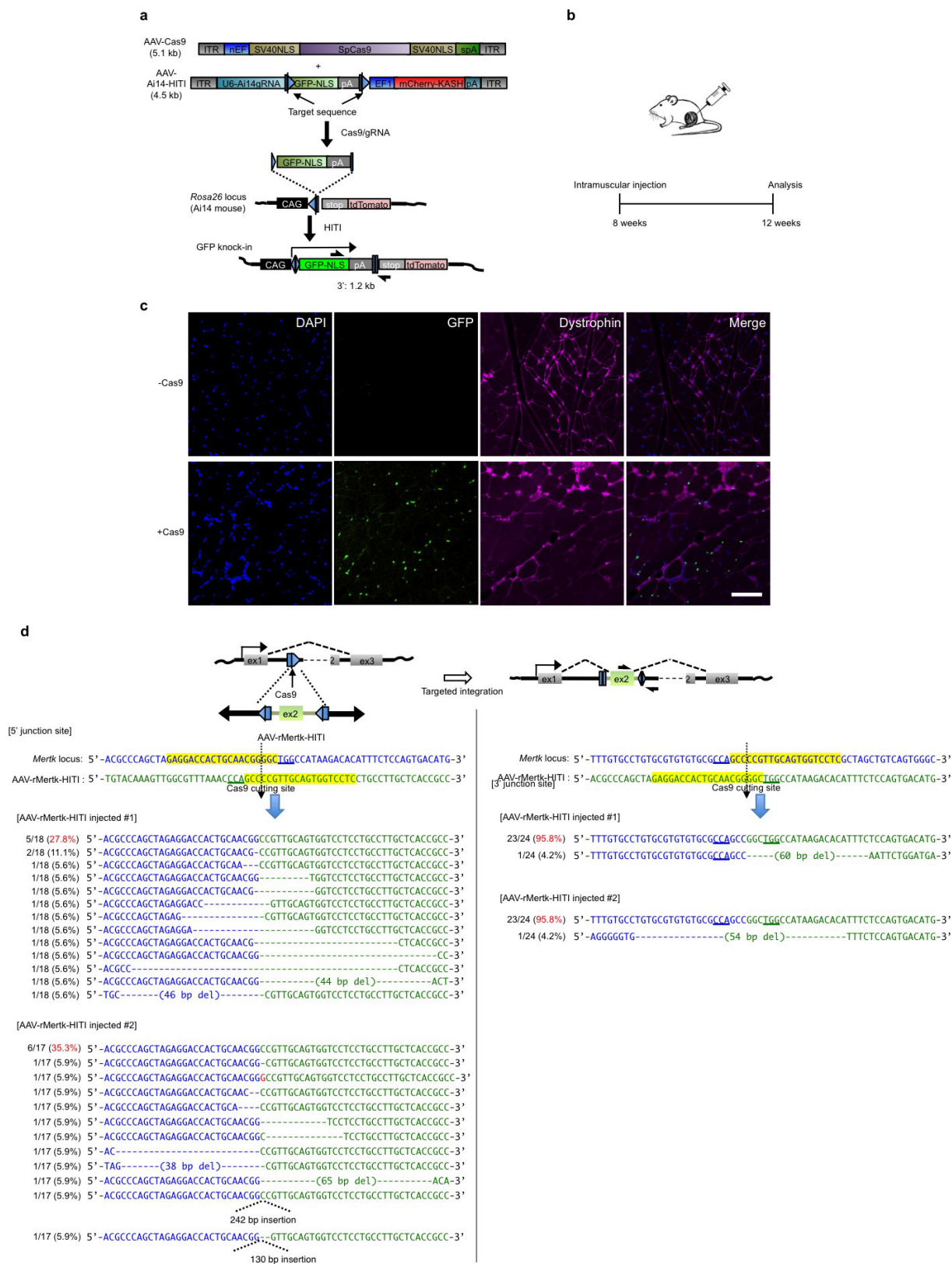
Ai14-GFPNLS-MC). Insets, higher magnification images. Scale bar, 100 μ m.

d, *In vivo* imaging of luciferase signals at days 7 and 14 after pressure-mediated kidney transfection of luciferase HITI constructs. Left mouse (–Cas9) was delivered with empty plasmid, Ai14gRNA-mCherry and Ai14-luc-MC. Right mouse (+Cas9) was delivered with CAG-Cas9, Ai14gRNA-mCherry and Ai14-luc-MC. Top, control wild-type mouse. Bottom, Ai14 mouse. **e**, *Ex vivo* luciferase imaging of stomach and oesophagus (St+E), heart (H), liver (Li), spleen (Sp), lungs (Lu), right (R) and left (L) kidney (K), pancreas (Pa), small intestine (SI), caecum (Ce) and colon (Co). Arrow shows luciferase signal in the right kidney. Top, wild-type mouse. Bottom, Ai14 mouse. **f**, Representative immunofluorescence images of GFP expression after electroporation of GFP-NLS HITI constructs into Ai14 mouse kidney. Top panel, donor-cut only control (CAG-Cas9, Scramble-gRNA-mCherry and Ai14-GFPNLS-MC-scramble). Middle panel, no Cas9 control (empty plasmid, Ai14gRNA-mCherry and Ai14-GFPNLS-MC). Bottom panel, GFP-NLS HITI (CAG-Cas9, Ai14gRNA-mCherry and Ai14-GFPNLS-MC). Scale bar, 100 μ m.



Extended Data Figure 7 | AAV8-mediated HITI in cultured mouse primary neurons. a, Representative immunofluorescence images of neurons infected with AAV-Cas9 and AAV-mTubb3. AAVs were packaged with serotype 8. Insets, higher magnification images. Scale bar, 50 μ m. **b**, Representative immunofluorescence images of GFP knock-in at the *Tubb3* locus in pan neurons after AAV8 infections. Top panel, AAV-mTubb3 only. Bottom panel, AAV-Cas9 and AAV-mTubb3. Scale bar, 100 μ m. The absolute GFP knock-in efficiency was shown in the upper

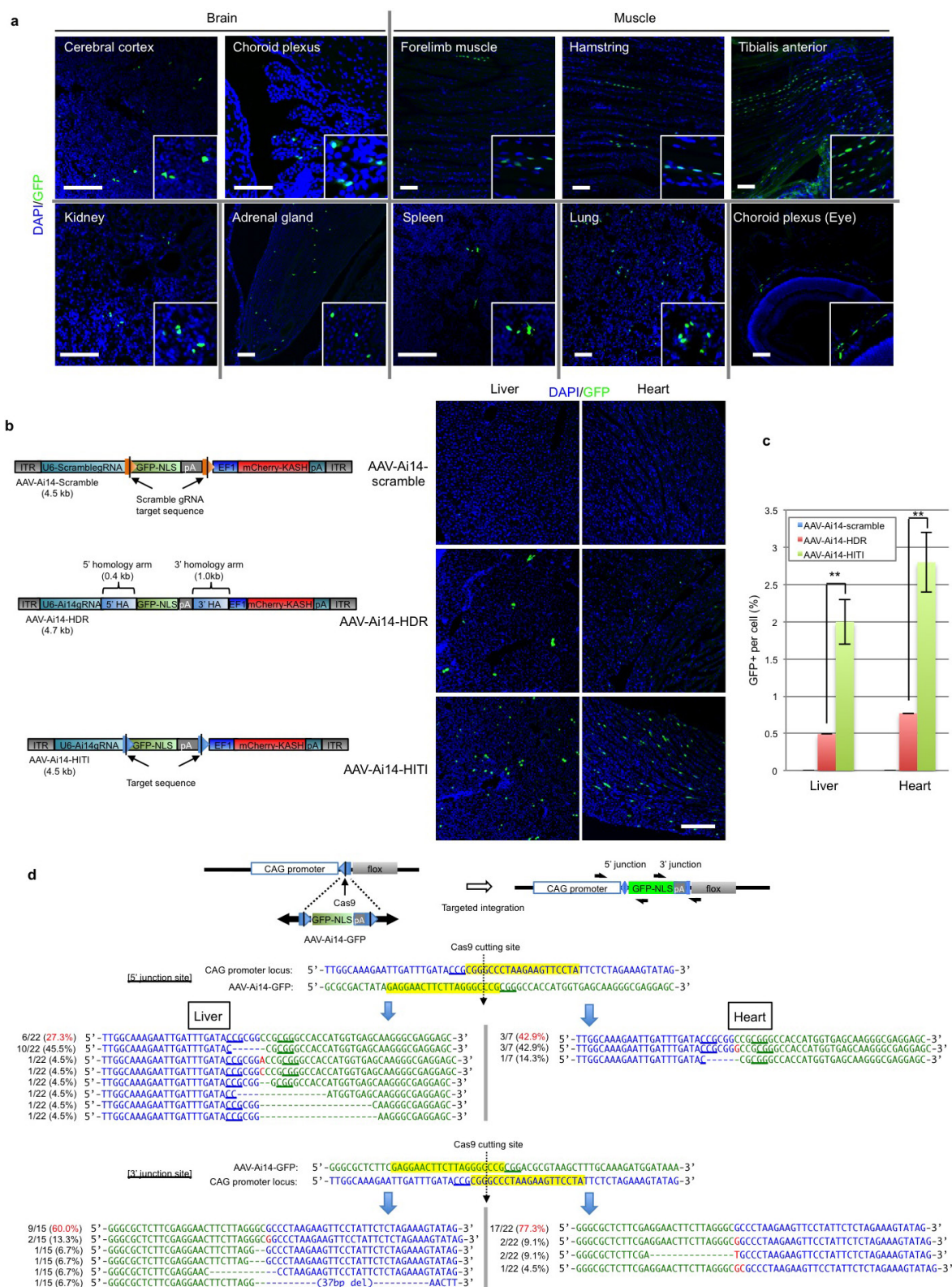
right corner of the picture. Results were obtained from three technical replicates and presented as mean \pm s.d. **c**, Intracellular distribution of the GFP expression after AAV8 infection (AAV-Cas9 and AAV-mTubb3). *n*, number of analysed cells. **d**, Validation of correct gene knock-in by PCR. **e**, Sequences of the 5' and 3' junction sites after GFP knock-in by HITI. For gel source image, see Supplementary Fig. 1. For source data, see Supplementary Table 10.

Extended Data Figure 8 | HITI via *in vivo* local injection of AAVs.

a, Schematic of AAV vectors for inserting GFP-NLS downstream of the CAG promoter at the Ail4 *Rosa26* locus. The AAVs were packaged with serotype 8. Half-arrows indicate PCR primers to validate correct gene knock-in. b, Schematic of the experimental design for *in vivo* targeted GFP-NLS gene knock-in by HITI via intramuscular (IM) injection in Ail4 mice. AAV8s (AAV-Cas9 and AAV-Ail4-HITI) were

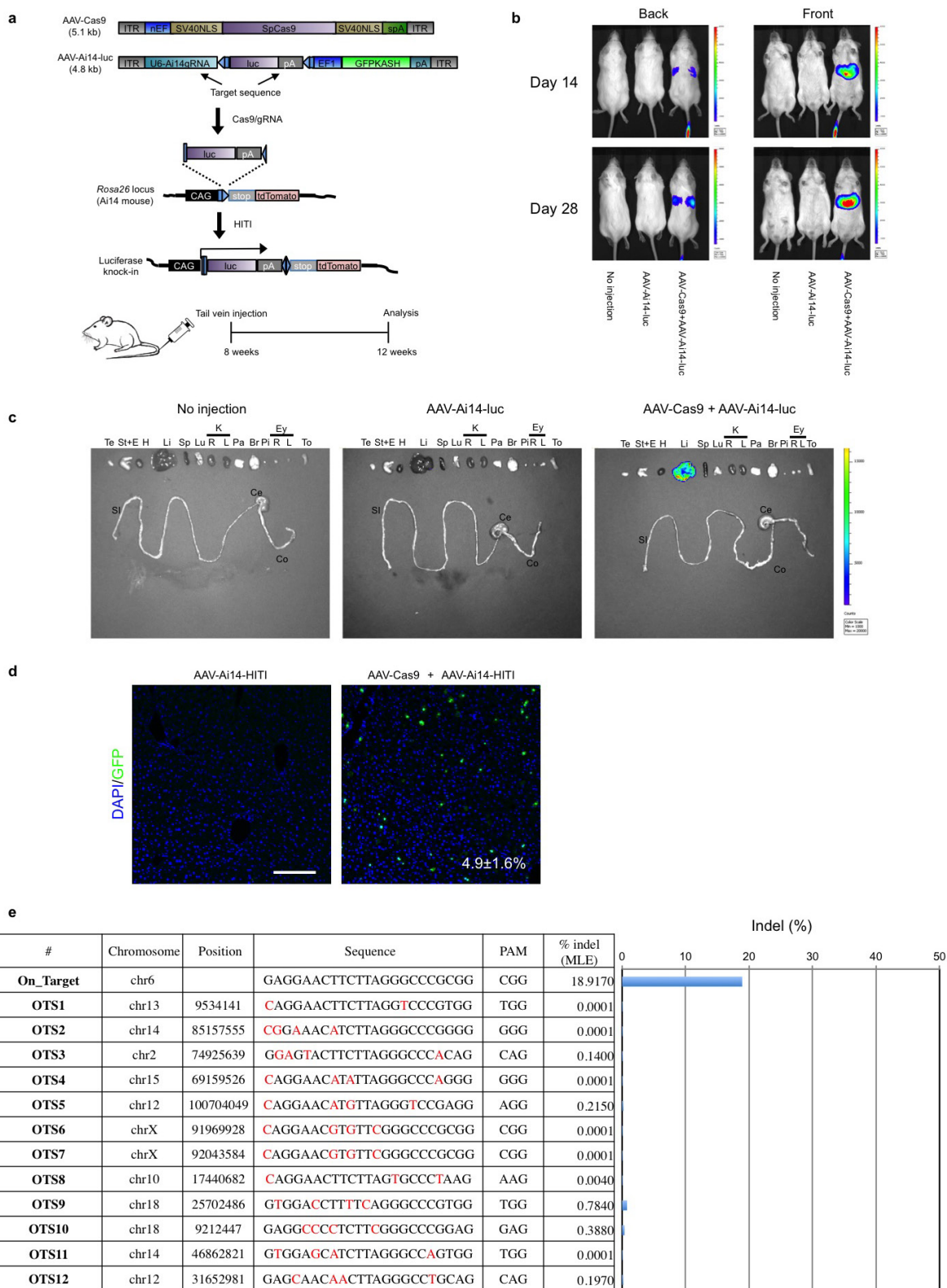
injected into quadriceps at 8 postnatal weeks and analysed at 12 weeks.

c, Immunofluorescence analysis of GFP expression after IM delivery of AAV8s. Top panel, no Cas9 control (AAV-Ai14-HITI only). Bottom panel, GFP-NLS HITI (AAV-Cas9 and AAV-Ai14-HITI). Dystrophin was used as a marker for muscle cytoskeletal protein. Scale bar, 100 μ m. d, Schematic and sequencing analyses of 5' and 3' junctions of the integration sites for *Mertk* exon 2 inserted by HITI in the eyes of RCS rats.



Extended Data Figure 9 | HITI-mediated GFP-NLS knock-in via systemic intravenous injection in neonatal Ai14 mice. **a**, Representative immunofluorescence images of GFP expression in the brain, muscle, kidney, adrenal gland, spleen, lung and choroid plexus of the eye after intravenous injection of AAV-Cas9 and AAV-Ai14-HITI. AAVs were packaged with serotype 8. Insets, higher magnification images. Scale bar, 100 μ m. **b**, Comparison of HDR- and HITI-mediated targeted gene knock-in via systemic intravenous injection in neonatal mice. Left panel shows a schematic of AAV vectors for knock-in GFP downstream of the CAG promoter at the Ai14 *Rosa26* locus. AAV-Ai14-scramble was used

as a donor cut-only control. The HDR donor (AAV-Ai14-HDR) has homology arms at both ends of the GFP-NLS-pA cassette. The AAVs were packaged with serotype 8 and co-infected with AAV-Cas9 via IV, same as Fig. 4a. **c**, Absolute knock-in efficiency measured by the percentage of GFP⁺ cells in the liver and heart by HDR or HITI. Results were obtained from five slides and presented as mean \pm s.d. ** P < 0.01, unpaired Student's *t*-test. **d**, Sequencing analyses of the 5' and 3' junction sites of heart and liver cells after GFP-NLS knock-in by HITI via intravenous AAV injections. For source data, see Supplementary Table 11.



Extended Data Figure 10 | HITI via *in vivo* systemic injection of AAVs in mice. **a**, Schematic of AAVs for knock-in luciferase downstream of the CAG promoter at the Ai14 *Rosa26* locus. AAVs (AAV-Cas9 and AAV-Ai14-luc) were systemically delivered via tail vein injection in 8-week-old Ai14 mice and analysed at 12 weeks. The AAVs were packaged with serotype 9. **b**, *In vivo* imaging of luciferase signals at days 14 and 28 post-tail vein injection of luciferase HITI constructs. **c**, *Ex vivo* luciferase imaging analysis of testis (Te), stomach and oesophagus (St+E), heart (H), liver (Li), spleen (Sp), lungs (Lu), right (R) and left (L) kidney (K), pancreas (Pa), brain (Br), pituitary (Pi), right (R) and left (L) eye (Ey),

Tongue (To), small intestine (SI), caecum (Ce) and colon (Co). **d**, Representative immunofluorescence images of GFP expression in the liver after tail vein injection of HITI GFP-NLS gene knock-in AAV9. The absolute efficiency of GFP knock-in was shown in the bottom right corner. Results were obtained from five slides and presented as mean \pm s.d. Scale bar, 200 μ m. **e**, A list of on- and off-target sites that were used to determine the indel frequency of HITI mediated genome modifications using genomic DNA isolated from the liver. The nucleotide letters shown in red are the individual mismatches in predicted off-target sites.

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature20108

Corrigendum: Design of a hyperstable 60-subunit protein icosahedron

Yang Hsia, Jacob B. Bale, Shane Gonen, Dan Shi, William Sheffler, Kimberly K. Fong, Una Nattermann, Chunfu Xu, Po-Ssu Huang, Rashmi Ravichandran, Sue Yi, Trisha N. Davis, Tamir Gonen, Neil P. King & David Baker

Nature **535**, 136–139 (2016); doi:10.1038/nature18010

In this Letter, the 60-subunit protein assembly presented more strongly resembles a wireframe dodecahedron than an icosahedron. We thank the reader who drew this to our attention and we agree that all instances of the word “icosahedron” should be changed to “dodecahedron”. In addition, “20 pentameric faces” in the boldface first paragraph should have read “12 pentagonal faces”, and throughout the text “pentameric faces” should have read “pentagonal faces”. The original Letter has not been corrected.

CORRIGENDUM

doi:10.1038/nature20144

Corrigendum: Fumarate is an epigenetic modifier that elicits epithelial-to-mesenchymal transition

Marco Sciacovelli, Emanuel Gonçalves,
Timothy Isaac Johnson, Vincent Roberto Zecchini,
Ana Sofia Henriques da Costa, Edoardo Gaude,
Alizee Vercauteren Drubbel, Sebastian Julian Theobald,
Sandra Riekje Abbo, Maxine Gia Binh Tran, Vinothini Rajeeve,
Simone Cardaci, Sarah Foster, Haiyang Yun, Pedro Cutillas,
Anne Warren, Vincent Gnanapragasam, Eyal Gottlieb,
Kristian Franze, Brian Huntly, Eamonn Richard Maher,
Patrick Henry Maxwell, Julio Saez-Rodriguez &
Christian Frezza

Nature **537**, 544–547 (2016); doi:10.1038/nature19353

In this Letter, the ArrayExpress accession number provided for the gene expression data for *Sdhb*-deficient cells should have been 'E-MTAB-4349', rather than the Affymetrix GeneChip platform accession 'A-AFFY-130'; this has been corrected in the online versions of the paper.

ERRATUM

doi:10.1038/nature19851

Erratum: Pancreatic stellate cells support tumour metabolism through autophagic alanine secretion

Cristovão M. Sousa, Douglas E. Biancur, Xiaoxu Wang, Christopher J. Halbrook, Mara H. Sherman, Li Zhang, Daniel Kremer, Rosa F. Hwang, Agnes K. Witkiewicz, Haoqiang Ying, John M. Asara, Ronald M. Evans, Lewis C. Cantley, Costas A. Lyssiotis & Alec C. Kimmelman

Nature **536**, 479–483 (2016); doi:10.1038/nature19084

In this Letter, during the production process, one of the images in Fig. 3a was inadvertently labelled with the incorrect PDAC cell line name. The second picture from the left should have been labelled “hPSC-LC3 + 8988T”, rather than “hPSC-LC3 + Tu8902”, to match the quantification provided in Fig. 3b. This has been corrected in the online versions of the manuscript.

CORRECTIONS & AMENDMENTS

ERRATUM

doi:10.1038/nature20106

Erratum: Pore-forming activity and structural autoinhibition of the gasdermin family

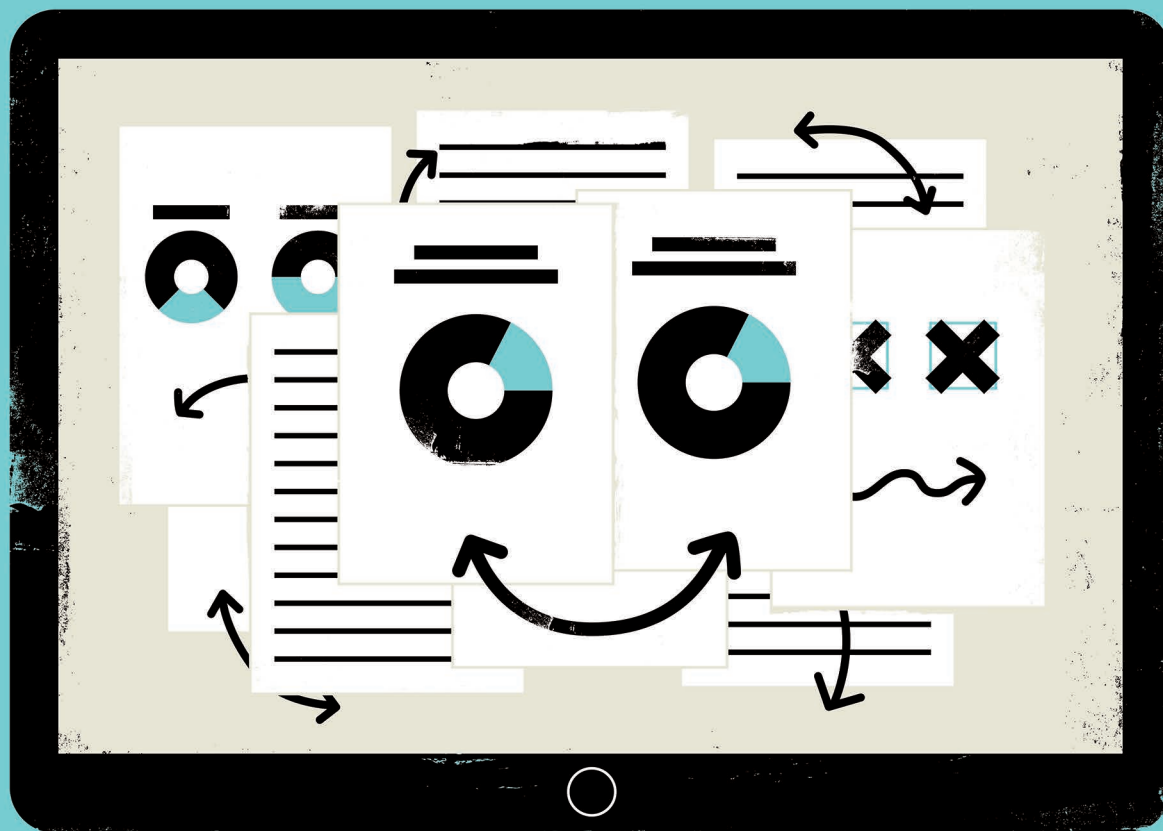
Jingjin Ding, Kun Wang, Wang Liu, Yang She, Qi Sun,
Jianjin Shi, Hanzi Sun, Da-Cheng Wang & Feng Shao

Nature **535**, 111–116 (2016); doi:10.1038/nature18590

In this Article, owing to a typesetter error the ‘received date’ was incorrectly shown as ‘6 March 2015’ instead of ‘6 March 2016’; this has been corrected in the online versions of the paper.

A SPELLCHECKER FOR STATISTICS

ILLUSTRATION BY THE PROJECT TWINS



Nuijten did not alert the papers' authors. But

Others foresee a positive change in the culture. Hartgerink and Nuijten have each received awards from organizations promoting open science. And in a PubPeer comment ▶

► on the original statcheck paper, psychology researcher Nick Brown of the University of Groningen in the Netherlands wrote that science might benefit if researchers stopped assuming that posts on the forum indicated that there was “something naughty” in a paper, and instead thought, “There’s a note on PubPeer, I will read it and evaluate it like a scientist.”

An automated tool makes researchers more likely to double-check their work, which is good for psychology, argues Simine Vazire, who studies self-perception at the University of California, Davis. “It will catch mistakes, but even more importantly it will make us more careful.”

That seems to appeal. Several thousand people have downloaded the free statcheck program, which works in the programming language R, or visited the web-based statcheck.io, which requires no programming knowledge. (Researchers who want to check selected results rather than whole papers can use online calculators such as ShinyApps.)

TECHNICAL CHECK

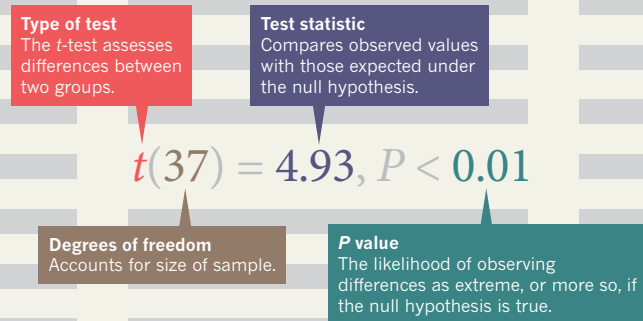
Most psychology papers report statistical tests in a standardized format, with related parameters that can be checked for inconsistencies. Statcheck — which so far works only for papers in this format — identifies and inspects a few common tests that calculate *P* values, a measure of how likely results are to arise by chance if, for instance, no real difference exists between two groups (see ‘What statcheck looks for’). Although statisticians have warned against it, a *P* value below 0.05 is often used as an arbitrary determiner of ‘statistical significance’, allowing results to be taken seriously and published.

Most of the errors that statcheck catches seem to be typos or copy-and-pasting mistakes, says Daniel Lakens, a cognitive psychologist at Eindhoven University of Technology in the Netherlands. After reading the statcheck paper, he decided to analyse the errors it reported that changed a result’s statistical significance. He found three main categories. Often, a researcher had inserted an incorrect sign, such as $P < 0.05$ instead of $P = 0.05$. In other cases, the calculations were set up to detect only particular relationships, such as positive or negative correlation, which was not always made explicit. Optimistic rounding was also common: *P* values of 0.055 reported as $P \leq 0.05$ made up 10% of detected errors that changed statistical significance, a rate that Lakens calls depressingly high.

But statcheck itself makes errors, says Thomas Schmidt, an experimental psychologist at the University of Kaiserslautern in Germany, who wrote a critique of the program (T. Schmidt Preprint at <http://arxiv.org/abs/1610.01010>; 2016) after it flagged two of

WHAT STATCHECK LOOKS FOR

This computer algorithm scans papers for statistical tests, uses reported results to recompute the *P* value and flags up inconsistencies.



his papers. For example, it does not always recognize necessary statistical adjustments.

When statcheck does detect an error, it cannot distinguish whether it is the *P* value or a related parameter that is incorrect. Schmidt says that, across the two of his papers that it scanned, statcheck failed to detect 43 *P* values, checked 137 and noted 35 “potentially incorrect statistical results”. Of those, 2 reflected *P*-value errors that did not change significance, 3 reflected errors in other parameters but did not affect *P* values, and 30 were improperly flagged.

Nuijten admits that statcheck can sometimes misidentify tests and overlook adjusted *P* values, but she notes that, in her original paper, it found similar rates of error to manual checks.

Nuijten and Hartgerink have been working hard, mostly successfully, to keep conversations amiable. Nuijten has posted detailed explanations about how statcheck works, with smiley emoji and friendly exclamation marks. Hartgerink is updating PubPeer posts with an improved version of the software. Both note that anyone can add comments on PubPeer to explain statcheck’s results, and that the posts state that results are not definitive. “The one thing I try to repeat over and over is that statcheck is automated software that will never be as accurate as a manual check,” says Nuijten.

Much of what statcheck flags up is trivial, but when authors do not respond, matters are left unresolved, says Elkan Akyürek, a psychologist at the University of Groningen. “Content-based discussion is getting a bit flooded.” Thought leaders such as neuropsychologist Dorothy Bishop of the University of Oxford, UK, worry that posts could distract from more serious discussions, or alienate people and make them less receptive to efforts to improve reproducibility. Heiko Hecht, a psychologist at Johannes Gutenberg University in Mainz, Germany, thinks it might have the opposite effect: “The program is still very immature, but in the long run could keep scientists honest.” Besides, he adds, if researchers made raw data available, anyone could check the results.

Some authors have expressed gratitude

for a chance to correct mistakes, although several have said that they should have the chance to review posts before they are made public. At least three have responded on PubPeer to explain errors. Two of them told *Nature* that the errors were typos that did not affect *P* values and were too trivial to justify a formal correction. As for Vazire, she hopes that automated reports will help researchers to get used to post-publication commentary. “I think it will help desensitize us to criticism,” she says.

EDITOR’S HELPER

In July this year, the journal *Psychological Science* began

running statcheck on submissions that got favourable first reviews, and discussing flagged inconsistencies with the authors. “I thought there might be some blowback or resistance,” says editor-in-chief Stephen Lindsay. “Reaction has been almost non-existent.” Of the few dozen runs so far, none of the errors has been egregious, he says, although there have been at least two instances in which authors have reported a *P* value as 0.05 when it was 0.054.

Lindsay says that statcheck reports are too confusing to share with authors directly. (For example, the program flags potential errors with the word TRUE.) Nuijten says that an upcoming version will be much more comprehensible to non-programmers. Meanwhile, she says, her team has been talking to publishers Elsevier and PLOS about adopting the program at their titles. And statcheck may soon have company: a more-comprehensive commercial program called StatReviewer is under development by other researchers. It is designed to analyse papers from a variety of fields, not just to double-check calculations but also to ensure that reporting requirements are followed.

Lindsay hopes that statcheck’s utility will fade over time as researchers stop manually entering statistical outcomes into their manuscripts; instead, the values would be directly inserted by the programs that produced them, and linked to their scripts. “The methodological leaders are using things like R markdown,” he says.

As for Schmidt, he thinks that statcheck could be useful in manuscript preparation, but it is not for beginners. “The greatest risk during prepublication is that unsophisticated users overestimate the program, relying blindly on its output,” Lakens is sticking to a manual system: one author of a paper does the analyses, and another checks them. That can detect errors that statcheck will not, such as transposing results.

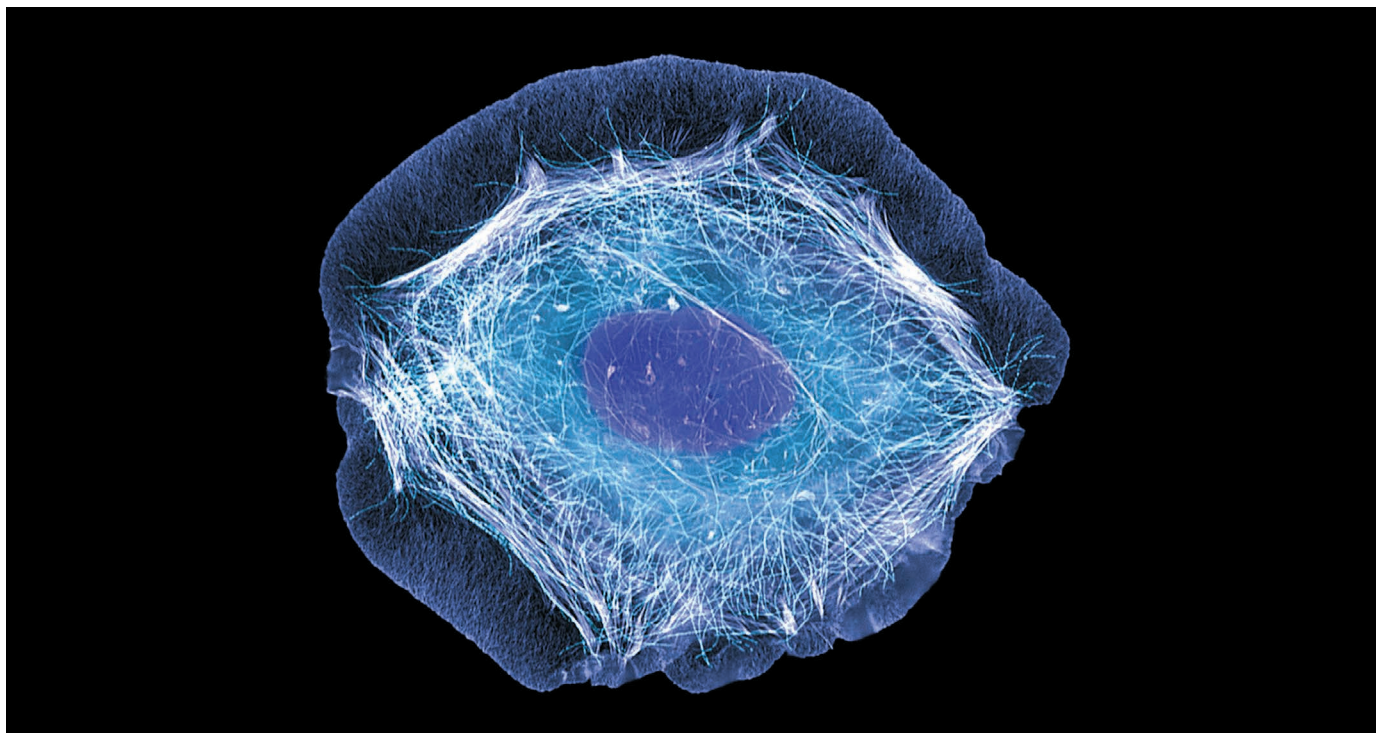
That approach makes sense to Nuijten. Her goal was never to fix statistical analysis. Statcheck is more like a standard spellchecker, she says: “a handy tool that sometimes says stupid things”. People laugh at the absurdities, but still use the tool to correct mistakes. ■

TECHNOLOGY FEATURE

METABOLOMICS: SMALL MOLECULES, SINGLE CELLS

Sensitive mass spectrometry and innovative cell-sampling techniques allow researchers to profile metabolites in single cells, but the field is still in its infancy.

TORSTEN WITTMANN/SPL



Researchers are developing innovative methods to understand the inner workings of individual cells.

BY MARISSA FESSENDEN

Sitting in his first-floor office, in an industrial lab space that opens onto a field of grazing cattle, analytical chemist Renato Zenobi explains one of the fundamental problems facing today's cell biologists. He traces out a curve representing the average concentration of a molecule in a theoretical cell population — a simple bell-shaped distribution. That distribution, he explains, can obscure complexity. To prove the point, he sketches two curves overlapping either side of the single peak, each representing a distinct phenotype in the population — and also consistent with that bell-shaped curve. “To really figure out if the distribution is multimodal or bimodal, you need to go down to the single-cell level,” says Zenobi, at the Swiss Federal Institute of Technology (ETH) in Zurich.

Cell heterogeneity is why some bacteria in

a clonal population can develop antibacterial resistance. It also gives rise to the different cell subpopulations in the brain. And it explains tumour relapses. The tools to detect those differences are only just becoming available.

“Recent technological advancements — especially those made just in the past two years — have revealed that individual cells within the same population may differ dramatically,” says Ananda Roy, programme leader of the National Institutes of Health (NIH) Common Fund working group for single-cell analysis in Bethesda, Maryland. “And these differences can have important consequences for health and disease.” Funders worldwide have queued up to support single-cell research. The NIH has invested in special initiatives to support single-cell profiling, starting with US\$2 million in 2014, and almost 60 groups have now received awards under the programme. A collaboration

between universities and companies in Japan launched the Society for Single-Cell Surveyor, which awards grants and holds symposia on single-cell analysis and technologies. And in October, experts discussed launching the International Human Cell Atlas Initiative, which aims to chart every type of human cell and its properties — an ambitious task that relies heavily on single-cell analysis.

Much of this work focuses on revealing cell-to-cell differences at the DNA level. Yet profiles of genes and epigenetic modifications merely outline a cell's potential. The rapid, dynamic responses a living cell has to its environment are better reflected in the metabolic transformations — and the resulting patterns of small molecules — that keep the cell powered, cycling and communicating with others.

“The metabolome is most directly related to the phenotype,” says Caroline Johnson, an ►

► analytical chemist at Yale School of Public Health in New Haven, Connecticut. It reveals the products of the genome and its protein output, as well as metabolites from diet, drugs and toxic compounds.

Yet its complexity means that metabolomics has lagged behind other 'omics'. Unlike DNA and RNA, metabolites cannot be amplified. Although some metabolites can reach millimolar concentrations, a single cell offers a limited volume for analysis. Exquisitely sensitive methods are needed to detect rarer compounds — and the abundances can change in seconds in a living cell. Metabolites also have a bewildering variety of structures. The Human Metabolome Database contains records for more than 42,000 metabolites, from sugars to peptides to cofactors. But the total may be significantly higher, and single analytical methods often struggle to capture the chemical diversity.

Still, the field is advancing, thanks to leaps in detection capabilities, increasingly sophisticated ways of isolating and handling single cells and developments in bioinformatics. "We are getting close to making single-cell metabolomics robust," says Jonathan Sweedler, an analytical chemist at the University of Illinois at Urbana-Champaign. "I can't see exactly how it is going to work, but I look at the increase in detectability and throughput of mass spectrometry and I say it will happen."

PUSHING SENSITIVITY

Profiling individual yeast cells allowed Zenobi's group to spot two phenotypes lurking in a genetically identical sample — one characterized by low levels of a metabolite called fructose 1,6-bisphosphate, and another with high levels¹. The difference probably comes down to different glucose-utilization strategies. The insight

doesn't have immediate biomedical applications, Zenobi admits, but it does illuminate a fundamental way in which cells work.

To draw such insights, Zenobi's team uses sophisticated techniques to isolate cells and to boost the sensitivity of its analytical approaches. Researchers generally use either mass spectrometry or nuclear magnetic resonance (NMR) to drive metabolomics studies. But because NMR is less sensitive, mass spectrometry has emerged as the method of choice, and there are many variations to improve its detection abilities, throughput or simply make it easier to pull the contents out of a cell. "The whole arsenal of mass spectrometry has been thrown at this problem," Zenobi says.

Mass spectrometry involves ionizing a sample to lend a charge to its constituent molecules. The charge then means that magnetic plates can nudge the molecules as they fly through a vacuum. Each molecule deflects by a different amount, depending on its mass-to-charge ratio, such that by the time the molecules reach the detector, they have resolved into their component parts. The data appear as a complicated plot of unidentified peaks, each corresponding to a different molecular entity.

The method is straightforward — except when applied to single cells. Attempting to detect just a few molecules in a vanishingly small volume pushes the limits of modern instrumentation.

Zenobi uses a specialized silicon slide to individually deliver hundreds of single cells into the mass spectrometer. To the naked eye, the slides seem to be covered with a fine black mesh. The mesh is a coating of a polymer called polysilazane, which has been laser micro-machined to create hundreds to thousands of reservoirs, each a couple of hundred

micrometres in diameter. When the researchers add a dilute solution of cells to the slide, the coating's repellent properties ensure that some liquid, and one or two cells, end up in each reservoir. The researchers then direct the spectrometer's laser to target each well in turn. One configuration of these slides, called microarrays for mass spectrometry (MAMS), is available from MilliporeSigma in St. Louis, Missouri.

Across the hall from Zenobi's office, graduate student Robert Steinhoff demonstrates how the slide fits into a mass spectrometer. The researchers use matrix-assisted laser desorption/ionization (MALDI) coupled to a time-of-flight analyser, which requires them to spot their slides with a chemical matrix to drive ionization. By using a matrix that minimizes interference with the signal given off by small molecules,

"Individual cells behave very interestingly and unexpectedly."

the team can detect metabolites in the low attomole (10^{-18} moles) range — and do so for about 1,000 cells per chip, which is relatively high

throughput in the single-cell world.

Steinhoff is also working to alter the substrate used in his reservoirs. "Making nanostructures — pillars — within the spot helps us see things more reliably," he explains. The mechanisms behind this effect aren't yet clear, but Steinhoff says that the cells end up clinging to the tops of the silicon columns.

Sweedler has developed a high-throughput approach that uses a computer to guide the laser to individual cells spread across a slide. His team can process about 10,000 cells per slide in this way. But for a more comprehensive look at the metabolome, Sweedler takes it one cell at a time. He uses a modified patch-clamp tool, which typically records cellular electrical signals, to withdraw roughly three picolitres of cytoplasm (about 10–40% of the total volume) from individual brain cells and deliver it into a mass spectrometer.

The limited throughput restricts analysis to a few dozen cells per experiment. Still, Sweedler's group has used it to detect about 60 metabolites from 30 neurons and astrocytes in slices of rat brain². The team focused on neurochemicals such as glutamate, but also detected amino acids and derivatives of ATP, among others. From that, they compiled unique profiles for the different cell types, providing a window into the cell-to-cell heterogeneity that makes the brain so complex, Sweedler says.

DIFFERENT STROKES FOR DIFFERENT SIZES

When dealing with single cells, size matters. Plant cells can be anywhere from 10 to 100 micrometres across. Mammalian cells tend to be smaller, on the scale of 10–20 micrometres. Microbial cells are smaller still, reaching down to the submicrometre range. As cell size varies, so too does the volume and thus the absolute number of metabolites. "It is obvious from an

SELECTED METABOLOMICS SOFTWARE

Name	Description
Human Metabolome Database	Free database of metabolites in the human body, including details relevant to chemical profiles, molecular biology and clinical chemistry. Links to pathway and structure-viewing tools and to other databases.
KEGG Pathway	Visualization tool linked to a collection of manually drawn pathway maps (KEGG database) with ability to locate and colour specific metabolites and pathways.
LipidBlast	A tandem-mass-spectral database for lipid identification.
MapMan	Tool for converting large data sets, such as those from gene-expression arrays, into metabolic pathway diagrams. Focuses on plant metabolomics.
MetaboAnalyst	Complete workflow for metabolomics data processing and analysis for both mass spectrometry and nuclear magnetic resonance data with common statistical analyses and heat-map visualizations.
Metabolites Biological Role (MBRole)	A server that integrates chemical and biological annotations from a handful of databases to identify potentially important metabolites and other molecules.
MS-Dial	Downloadable program that spots and identifies peaks in mass-spectrometry data, developed for small molecules.
OpenMS	Open-source library for C++-based data analysis and management that includes infrastructure for software development. Mainly for processing raw data.
XCMS Online	Cloud-based platform for metabolite profiling from raw liquid-chromatography and mass-spectrometry data and for data sharing.
Yeast Metabolome Database	Manually curated collection of yeast metabolites and their chemical properties, with links to spectral and chemical databases.

analytical perspective there will be no single method that can deal with all these volumes,” says Akos Vertes, a chemist at George Washington University in Washington DC. His lab uses different methods for different cell sizes.

For the largest cells, the team uses a sharpened optical fibre to deliver infrared light directly into the cell. The light excites the oxygen–hydrogen bonds in the water of a cell, causing the cell to explode and eject its contents. The escaping material then meets an aerosolized, ionized liquid called an electrospray, which charges the molecules for mass spectrometry. The advantage of this technique is that single cells can be profiled while still embedded in tissue. But it also can be very slow, because each cell generally needs to be poked with the fibre by “a very patient graduate student”, Vertes says. He recently automated the process, using a computer to manoeuvre the sample stage.

The smallest cells are deposited on a nanopillar-covered surface, also made of silicon although fabricated differently from Steinhoff’s. Imaging the entire surface reveals where the instrument’s ion beam needs to aim to target single cells. With this method, the team can routinely detect metabolites at femtomole (10^{-15} moles) levels³. But the investigators estimate their lower limit of detection at about 800 zeptomoles (one zeptomole is 10^{-21} of a mole), or about 482,000 molecules.

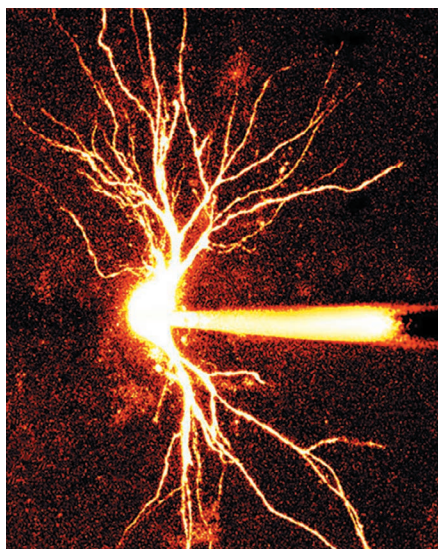
It is possible to go even smaller. Bioanalytical chemist Andrew Ewing at the University of Gothenburg in Sweden examines the small-molecule content of neural vesicles. These vesicles are key to delivering and releasing chemicals into the space between cells to facilitate cell-to-cell communication. “It’s tricky because you get very little signal with so few molecules in there,” he says.

Ewing uses an approach called nanoscale secondary ion mass spectrometry (NanoSIMS), which bombards a sample surface with a high-energy beam of caesium ions. This assault ejects charged particles from the surface, and they can then be analysed in a mass spectrometer to determine the material’s composition. Ewing’s group uses the method to assess distribution of the neurotransmitter dopamine. And by correlating the NanoSIMS data with transmission electron microscopy to observe a single vesicle as it loads and unloads dopamine, the researchers determined that a vesicle’s inner shape might regulate how quickly this process happens⁴.

SEEING IS BELIEVING

At the RIKEN Quantitative Biology Center in Osaka, Japan, chemist Tsutomu Masujima uses a video feed to target single cells for mass spectrometry.

“Individual cells behave very interestingly and unexpectedly, so I like to see [them] as much as possible,” Masujima says. His approach involves inserting a nanospray needle directly into a cell under video observation, sucking out the contents and then using the same needle to



The cytoplasm of a cell can now be extracted and its contents analysed.

inject the contents into the mass spectrometer. Adding the visual component allows his group to complete delicate manoeuvres, such as capturing and analysing the amino acid and lipid contents of single white blood cells and tumour cells circulating in the blood⁵.

But it also allows his team to estimate molecular abundance. Often, such a seemingly trivial calculation is complicated, because researchers aren’t sure exactly how much volume they are analysing. So Masujima and his colleagues use 3D microscopy to observe the cell before and after removing a small amount of cytoplasm. By measuring the resulting deformation, they can deduce the volume they removed as well as the location from which it was taken⁶. In one case, they sampled a cytosolic metabolite called methionine sulfoxide, determining that they had captured 5.9 zeptomoles of it.

MAKING SENSE OF THE DATA

Fortunately, says Gary Siuzdak, a chemist who heads the Scripps Center for Metabolomics and Mass Spectrometry in La Jolla, California, bioinformatics tools are available to make sense of such findings.

Siuzdak’s centre runs a cloud-based metabolomic analysis platform called XCMS Online (see ‘Selected metabolomics software’), whose 12,000-plus users have collectively shared more than 120,000 jobs. Few of those jobs have involved single cells, Siuzdak acknowledges, but that doesn’t mean that they are inherently incompatible with the software. “On the bioinformatics side, I don’t see a major issue with doing these experiments,” he says.

Rather, the main challenge with single-cell metabolomics is one of instrumentation: devices that can analyse enough single cells, and enough metabolites per cell, for the results to be statistically meaningful. “The primary issue with single cells is that the hardware still needs to be improved,” Siuzdak says. Studies tend

to profile tens or maybe a hundred different molecules. But a single yeast cell can have some 600 metabolites⁷. As a result, even the most sensitive analytical techniques are picking up only the easiest-to-detect, most prevalent molecules in a cell; less-common ones fall below the radar.

A possible solution to this problem, says computational biologist Jianguo Xia of McGill University in Montreal, Canada, may lie in population-based tools and in various ‘omics. “The pipelines and tools that have been developed for metabolomics data sets generated from bulk cell populations can be reused,” he says. All that may be needed are slight modifications to the data-processing and normalization methods that researchers use. “Single-cell transcriptomics is already established, and we can learn from it to accelerate bioinformatics development for single-cell metabolomics,” he adds.

Other researchers are pursuing single-cell strategies that are based on methods other than mass spectrometry — in particular, using living cells. Matthias Heinemann, for instance, a former postdoc in Zenobi’s lab, now heads a molecular systems-biology group at the University of Groningen in the Netherlands. There, he uses his background in analytical chemistry to “find other ways to zoom into single cells”. In one approach, his group uses fluorescent molecular sensors to quantify metabolites, such as ATP. In a forthcoming paper in *Molecular Cell*, his group has used time-lapse microscopy to watch levels of ATP and another metabolite, NADH (which is autofluorescent), oscillate as the cells go through the cell cycle⁸.

Key technical challenges have already been solved when it comes to single-cell metabolomics, Heinemann says. “What is needed now is to do tedious development and validation work.” The process might not be glamorous, he admits, but it is essential if single-cell metabolomics is ever to make the leap from proof of concept to answering fundamental biological questions.

“We are often so happy to detect unique molecules, but my point is: why?” Masujima says. “What is behind this finding, why does this molecule come up?” Without that insight, techniques run the risk of being mere gimmicks that fail to address questions of biological significance. “I don’t want to be a scientist who does that kind of science,” he says. ■

Marissa Fessenden is a freelance writer in Bozeman, Montana.

1. Ibáñez, A. J. et al. *Proc. Natl. Acad. Sci. USA* **110**, 8790–8794 (2013).
2. Aerts, J. T. et al. *Anal. Chem.* **86**, 3203–3208 (2014).
3. Walker, B. N., Antonakos, C., Retterer, S. T. & Vertes, A. *Angew. Chem. Int. Ed. Engl.* **52**, 3650–3653 (2013).
4. Lovrić, J. et al. *ACS Nano* (in the press).
5. Hiyyama, E. et al. *Anal. Sci.* **31**, 1215–1217 (2015).
6. Ali, A. et al. *Anal. Sci.* **32**, 125–127 (2016).
7. Förster, J., Famili, I., Fu, P., Palsson, B. Ø. & Nielsen, J. *Genome Res.* **13**, 244–253 (2003).
8. Papagiannakis, A., Niebel, B., Wit, E. & Heinemann, M. *Mol. Cell* <http://dx.doi.org/10.1016/j.molcel.2016.11.018> (in the press).

CAREERS

COLLABORATION Women have fewer co-authors on papers than do men **p.159**

EQUALITY UK gender pay gap in science and technology is narrowing **p.159**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



ADAPTED FROM GETTY

HUMAN BEHAVIOUR

Find your voice

Technology and practice can help shy and introverted researchers to succeed when reticence is risky.

BY JULIA ROSEN

Generally speaking, scientists aren't known as a gregarious bunch. Many identify as bookish, introverted, perhaps even a bit awkward. Yet those with more outgoing, extroverted traits might find it easier to thrive in today's scientific culture. That's because researchers in academia and

industry often have to step into the spotlight, by presenting their results at seminars and meetings and forging new relationships with colleagues, funders and, increasingly, the public.

Mastering these skills is especially important for young scientists who are trying to build their reputations and advance their careers. But for many shy or introverted researchers, these tasks can feel daunting, if not downright

terrifying. They can even cause some to question their place in science, says Louise Harkness, a postdoc at the Woolcock Institute of Medical Research in Sydney, Australia, who has blogged about the challenges of being an introverted scientist. "A future in academia is hard for the best scientists," says Harkness, who studies treatments for respiratory disorders. "Let alone for quiet scientists who are too shy to put their work forward."

Still, quiet scientists can compete successfully with their more loquacious counterparts by cultivating their public-speaking and networking skills, as well as by engaging in creative methods of self-promotion that fit their personalities. Researchers will need to acknowledge the political dimensions of professional science and examine their own personality traits and motivations to find approaches that work best for them.

RULES OF THE GAME

Along with the myth that all scientists are introverts, there is also a widespread perception that science operates as a pure meritocracy. Many young researchers think that they just need to do good research and the rest will follow, says Donna Dean, a retired administrator at the US National Institutes of Health and an executive consultant on leadership and talent development for the US Association of Women in Science. That's usually not the case, Dean says. "We can't just sit around and do nothing and assume that people will recognize our achievements."

Indeed, Jonathan Cheek, a personality psychologist at Wellesley College in Massachusetts, says that shy or introverted people can easily get overlooked in a culture of self-promotion. "Social communication skills, such as public speaking, are the largest predictor of career success outside of whatever the technical requirements for that career are," he says. That may not seem fair, he admits, but it's reality.

Acknowledging the importance of 'soft skills' is a good first step, Cheek says, particularly for certain types of introvert (Cheek and his colleagues recognize four different categories: social, thinking, anxious and restrained introversion). Not all introverts are shy, and some of them — all except those who are anxious introverts, according to Cheek — avoid speaking up and drawing attention to themselves simply because they don't wish to or don't find the behaviour rewarding. For those scientists, he says, it can be enough to recognize that there are tangible benefits to ►

► engaging in some form of self-promotion, even if it doesn't come naturally.

For others, the barriers are greater. People who experience general shyness feel discomfort when talking to strangers or in front of crowds (Cheek also helped to develop a shyness scale). And researchers who might sometimes feel unwelcome in science because of their identity — including women, minorities and those in the lesbian, gay, bisexual, transgender and queer (LGBTQ) community — can find themselves struggling to speak out in professional settings, Dean says. She adds that such discomfort might stem from a feeling that they bear the burden of representing their entire demographic group, or because they have been conditioned to be quiet as a result of their background.

"I came to this realization that if I don't get up there and present my results, the world is missing out."

Many in the scientific community agree on the need to help those researchers to amplify their voices, but in the meantime, researchers can help themselves by weighing the costs and benefits of staying quiet. "You have to think about, 'What's standing between me and my goals?'" says Cheek, who identifies as an ambivert, or someone who has both introverted and extroverted traits. If people have already invested years of their life in graduate studies, Cheek says, it's likely that they have a strong stake in continuing their scientific career and so will be willing to push past their shyness.

Sometimes, it just takes finding the right motivation, says Harkness, who overcame some of her quiet tendencies while doing her PhD research at the University of Sydney on gene regulation in asthmatic muscle cells. "I came to this realization that if I don't get up there and present my results, the world is missing out on these results and my thought process," she says (see 'Embrace your quietness').

Almost all scientists, at one point or another, have to share their research in front of crowds — a task that strikes dread into the hearts of many, not just the introverted and shy. Some surveys, such as the Chapman University Survey on American Fears conducted in 2014, show that in the United States, fear of public speaking often tops people's list of phobias, beating even fear of drowning in some cases. "Early in my PhD, I recognized it was something I was abysmal at," says Paul Brack, a PhD student at Loughborough University, UK, who studies ways to produce hydrogen for fuel cells. "I wanted to become average — that was my aspiration."

Fortunately, Cheek says, public speaking is not as hard to learn as many people fear, and doesn't require quiet researchers to become extroverts. The main reason, he adds, that most people loathe public speaking is that they haven't done it very much, and getting better at it just takes practice.

Many universities offer resources to help scientists to become comfortable presenting at meetings and to hone their speaking skills, says biochemist Kate Sleeth, interim associate dean of administration and student development at City of Hope hospital in Duarte, California. If an institution has no such offerings, Sleeth — also an introvert, who now chairs the board of directors of the National Postdoctoral Association — recommends seeking out groups such as Toastmasters International, a non-profit organization dedicated to helping its members to become better communicators.

Another strategy is for researchers to develop a presentation style that feels comfortable to them. For Harkness, that involved using her talks to illustrate her thought process, rather than just to disseminate her findings. "I want to take people through the story," she says. Stepping through the evolution of a research project actually made her feel better about presenting her work, she says. "I'm quite proud to show it."

Networking can also be adapted to individual preferences, despite the intimidating connotation that it has for many scientists. "The word 'networking' makes a lot of people feel like they are going to have to come up with some sort of beautifully flowing conversational piece," says Brack, who wrote about the subject in a *Naturejobs* blogpost last year (see go.nature.com/2fx60wc). But he has devised several ways to network that suit him, as an introvert who also used to be very shy.

One strategy involves approaching individuals — rather than big groups — at networking events at meetings with a question or two in mind ahead of time. If Brack strikes up a conversation with a fellow graduate student,

he usually leads with questions about their research, adviser and university. You don't even have to stick to science, says Dean, who still finds networking hard. Perhaps you notice something on someone's name badge — such as being from the same place — or share a hobby or other connection. "Get people talking about themselves," she says.

Dean advises young scientists to set a goal of talking to two or three new people each time they go to a conference, and she urges them to avoid describing their work in self-deprecating terms. Sleeth also suggests taking along an outgoing friend who will help you to feel comfortable, but who will not hog the limelight. Quiet scientists might also consider collaborating with more-extroverted colleagues on research. "It makes it so much easier," Sleeth says.

Ultimately, even if these tasks don't ever feel natural to many quiet scientists, those scientists should not despair, says Steve Blank, who teaches entrepreneurship at Stanford University in California and has worked with the US National Science Foundation's Innovation Corps Program, which helps scientists to commercialize their discoveries. "By definition, scientists are pretty smart," Blank says. "While you might not have it in your gut, you have enough computing power to emulate it."

PLAYING THE LONG GAME

When making big career choices, quiet scientists might want to consider how different paths in science might suit their personality. For instance, academia probably entails teaching and giving many public talks, whereas government agencies might require more lab work and meetings with agency managers. As entrepreneurs in the tech industry, Blank says, scientists have to sell their ideas to investors and customers. "If you want a leadership role, I'd say the biggest thing you need to learn is to communicate," he says. And that often involves at least learning to emulate an extrovert.

Because of the diverse demands of different scientific trajectories, Cheek recommends that early-career scientists look at literature such as Holland's theory of vocational choice, developed by the late John Holland, a psychologist at Johns Hopkins University in Baltimore, Maryland. "It's sort of a theory about how work environments have personalities," he says. Both people and occupations are ranked against Holland's framework, and three of the categories — realistic, investigative and artistic — are well suited to introverts. These might correspond to more applied, theoretical and creative career paths, respectively.

Scientists should not let such classifications dissuade them from following their aspirations, Cheek says, but they should consider whether their personality is compatible with their intended career choice. "Your preference, when it interacts with the structure of the field, maybe doesn't make your favourite thing the most rewarded thing," he says. Part of finding

SHY SUCCESS

Embrace your quietness

Try these tips for networking and promoting your work.

- Practise public speaking with colleagues or mentors or seek the help of a university programme or organization such as Toastmasters International.
- Develop a presentation style that feels natural to you, such as storytelling.
- At conferences and other professional gatherings, set a goal of talking to two or three new people.
- Use digital tools such as Twitter and Google Scholar to elevate your online profile.
- Join committees and seek out sponsors who can help you to make connections.
- Consider how different paths in science might fit with your personality traits and preferences. **J.R.**



Self-declared introvert Paul Brack presents at a 2016 Royal Society of Chemistry conference.

ROYAL SOC. CHEM./LOUGHBOROUGH UNIV.

a fulfilling career is about finding a good fit.

That's what Harkness has begun to realize, although she didn't formally consult Holland's theory. After her PhD, she moved to the Woolcock Institute, where research groups are smaller and more tightly knit, she says. That makes it a good place for her as a young, introverted scientist, she explains.

WRITING NEW RULES

Public speaking and in-person networking are seen as crucial to success in science and so many other fields partly because our culture tends to be geared towards extroverts. At least, that's the argument in *Quiet* (Broadway, 2013), a book by US writer and lecturer Susan Cain about the power of introverts. Her central argument is that society often treats introversion as a personality flaw, but that introverts should be valued. Quiet scientists can show off their strengths in a variety of ways.

David Steen has had great success with social media. A wildlife ecologist at Auburn University in Alabama, he is a proud introvert. "I don't manage it — I embrace it," he says. Steen describes himself as the guy who sits in the back of meetings and doesn't say much; similarly to many introverts, he prefers to gather his thoughts before he speaks.

Then, he communicates through tools such as Twitter, where he has more than 12,800 followers (*Slate* magazine crowned him "best biologist on Twitter" in 2015). The medium offers him the opportunity to gain visibility and to interact with the broader scientific community at his own pace and on his terms. "Can take all day to write a tweet," said Steen, in a tweet. "Compare that to the brief window of time you have during a

meeting to come up with verbal eloquence."

Cheek recommends creating a Google Scholar profile and becoming active on ResearchGate and Academia.edu. Researchers can also use LinkedIn to advertise scientific qualifications and promote publications and awards, Brack says. "You can share that in a way that you don't feel like you are ramming it in people's faces, or being overbearing."

Shy and introverted researchers can advance their careers in other ways that feel compatible with their personalities. Brack recommends joining committees, as he did recently to help organize a chemistry conference in Scotland. "I find that I need to be around someone quite a while before I'm comfortable enough to really talk to them and make that sort of a connection," he says. "Being in a committee is quite useful for that." Researchers can also follow up on meetings and seminars with e-mails to contribute to and stay involved in scientific discussions, particularly if they don't want to speak up in a group setting.

Regardless of the strategy that researchers choose, Sleeth recommends seeking out a mentor, or even a sponsor — whom she defines as someone who will advocate for and endorse young scientists — and especially shy and introverted ones. That person could be their adviser or it could be a co-author or colleague, she says. The important thing is that they help to open doors for young scientists and advertise their strengths and accomplishments. "Because then you're not bragging," she says. "Someone else is doing it on your behalf." ■

Julia Rosen is a freelance writer in Portland, Oregon.

GENDER

Co-author differences

Female faculty members at US research universities have fewer co-authors than do men over the entirety of their career, according to a study published last month (X. H. T. Zeng *et al. PLoS Biol.* **14**, e1002573; 2016). The study says that on average, women have shorter careers and lower publication rates, both of which contribute to the difference in their number of co-authors. Analysing the publication records of 3,980 faculty members across six disciplines in science, technology, engineering and mathematics, the authors found that women are more open than are men to new collaborations, a career strategy that has been found to produce higher-impact research. Yet the study also found gender differences in some disciplines. For example, female molecular biologists typically work in smaller teams than do their male counterparts. Looking specifically at the sub-field of genomics, which tends to produce work conducted by large teams, the authors also found that female scientists are under-represented as co-authors.

INCOME

UK pay gap

Men who work in the UK science and technology sector earn 24% more than women who work in the sector, according to the Annual Survey of Hours and Earnings from the UK Office for National Statistics (ONS). The survey found that for full- and part-time workers across all sectors, the gender pay gap is 18%, the lowest since the survey was launched in 1997, when the gap stood at 27.5%. The UK government will require all employers with more than 250 staff members to begin publishing their gender pay and bonus gaps from April 2017 in a bid to help women to overcome income barriers. In another effort to achieve parity, the government will coordinate with businesses to raise the number of women on executive boards to 33% by 2020. It has already extended the right to request flexible working, which could include flextime and teleworking, to all employees; introduced shared parental leave; offered support to female entrepreneurs for launching and growing a business; and increased the national living wage. Tackling the pay gap could add £150 billion (\$US187 billion) to the nation's annual gross domestic product in 2025, according to estimates.

THE FIRST FRAGMENTED CHURCH OF ENTROPY

Order from chaos.

BY STEVEN FISCHER

There's a sign hanging up on the side of the fence, although I'm not sure it really merits the word 'sign'. It looks more like a pasted preschool collage — letters snatched from holo displays and antique marquees mashed together like Mark's Thursday-night casserole.

I check the screen on my wrist, pulling up the infofeed and matching it to the words on the sign.

First Fragmented Church of Entropy

Speaking of Mark, where the heck is he? He's the one who dragged me here, after all.

I take another look at the building — crumbling brickwork and shattered windows — and step up to where the door should be. It's gone, of course, rusted off the hinges and lying in the weeds to the side of the walkway.

This is already worse than I'd expected.

I step inside and almost lose my footing. The floor's a jumble of tiles, polyfibre and something green and alive. Like someone flew a trash cruiser around town and then dumped whatever they found here.

The church is packed. It's just one big room littered with places to sit. I'd call them seats, but you know the drill by now. In one corner, a man and his son are smashing vases with a baseball bat. Across the room, someone in a devil costume stuffs little balls into one half of a clear container while his friend operates a door to the other half.

I catch sight of Mark lounging on half a couch beside a tree stump and hurry over to him, sidestepping a swimming pool that *obviously* belongs here. "Hi there!" I say, glad to see something familiar in a place so out of whack.

He kisses me and grins. "What do you think?"

"Well, it's certainly... something." I glance at the tangle of display screens, holo projectors and old neon bar signs hanging from the ceiling. "Although I imagine they're breaking a few building codes."

Mark smiles and shrugs. "Everything breaks, eventually."

He's been saying that a lot, lately. And although his suddenly carefree attitude has been a pleasant change, my PhD screams whenever I hear it. I open my mouth to explain that there's more to entropy than disorganized decorating schemes and breaking



things, but he's staring at the stage with a big, giddy grin.

"It's starting," he whispers.

The lights dim — well, the ones that work do — and the room explodes with noise. I would say music, if it wasn't so damn painful to hear.

I look around and realize the 'band' is scattered throughout the auditorium. There's a guy up on stage, sure enough, hammering away on an oversized electric xylophone, but there's also a lady lounging right behind us, revving the dismembered engine of a hoverbike in what she must know isn't a rhythm. I swear I hear a goose and half the rest of a farm somewhere, but I can't quite place them.

And then it stops.

A bespectacled old man gets up from his chair and shuffles on to the stage. Not exactly what I'd expected from a cult leader. I'm sorry, not a cult — Mark's not a fan of that word — 'a progressive, enlightened community'.

"Good morning," he says. And in the first normal event since I've arrived, everybody says it back. He smiles, takes the digital prompter in his hands, and throws it into the air.

"Everything breaks!" he proclaims. The prompter shatters on the floor, and everybody claps.

Aaaaaand, we're back to crazy.

He looks out at the crowd for a moment, his frazzled, unkempt hair threatening to blind him. "I hear we have a visitor in our congregation today."

I do my best to sink into the seat, but he points straight at me. "Welcome!" he says. "Why don't you come up front?"

Mark's practically giggling with excitement. He grabs my arm and tries to pull me to my feet. "Come on, Jan," he whispers. "It'll be fun."

I give him my best glare and work my way up to the stage. The pastor (priest? head thermodynamicist?) turns to me expectantly. "So, what brought you here today?"

I point at Mark. "He did." If he's making me be here, he better at least take the blame.

The man shakes his head. "No, no, no. What *really* brought you here?"

My mouth shuts. Because I'm worried my boyfriend signed on with some desperate freaks doesn't seem like a viable answer. "I suppose I have a few questions."

The room bursts into applause. For this?

"Well," old white-hair replies, "ask away."

I know what Mark would say. That I'm stressed out and harassed and constantly glued to a schedule and a screen. That his sudden ability to take life in his stride since he started coming here has me intrigued. And, it's true, but...

Is that — is that a bonfire by the back door?

Enough is enough. "Yeah. I was wondering how a church that claims to worship disorder justifies getting a group together each week, in one building for a set purpose, and teaching them about a topic discovered through centuries of highly regimented thought and experimentation. You can see the irony, right?"

I wait for him to say something, but he just stares. Then, he walks off the stage and right out the front door. One by one, the rest of the congregation leaves too, until it's just me and Mark in the auditorium.

I walk down the steps and take his hand as we head towards the door. "I'm really sorry about that," I mutter. "I know I promised to be civil."

Outside, the old man is tearing the sign down letter by letter.

Mark shrugs and grins back at me. "Everything breaks, eventually." ■

Steven Fischer is a student living in southern Wisconsin. In his free time, he enjoys dabbling in science fiction and exploring the great northwoods with his wife.

ILLUSTRATION BY JACEY



Produced with the support
of a grant from:



Reconnecting with
the nervous system

natureOUTLOOK

MULTIPLE SCLEROSIS

1 December 2016 / Vol 540 / Issue No 7631



Cover art: Señor Salme/
Synergy Art

Editorial

Herb Brody, Michelle Grayson, Nick Haines, Richard Hodson, Brian Owens, Jenny Rooke

Art & Design

Mohamed Ashour, Kate Duncan, Wesley Fernandes, Wojtek Urbanek

Production

Matthew Carey, Ian Pope, Karl Smart

Sponsorship

David Bagshaw, Samantha Meyer

Marketing

Nicole Jackson

Project Manager

Anastasia Panoutsou

Art Director

Kelly Buckheit Krause

Publisher

Richard Hughes

Editorial Director,

Partnership Media

Stephen Pincock

Chief Magazine Editor

Helen Pearson

Editor-in-Chief

Philip Campbell

Multiple sclerosis is a devastating disease that induces the body's own immune system to eat away at the central nervous system, slowly robbing patients of their physical mobility. It is also mysterious. Despite years of research, the cause remains elusive, and treatments are few and far between.

Progress is slow, but measurable. The study of clusters of cases in Canada is starting to reveal the complicated interactions between genetics, geography and a common viral infection that may lie behind the disease (see page S4). And new treatments, many based on existing drugs, are starting to show promise for the more debilitating progressive form of the disease, which has until recently been largely ignored and forgotten (S7). Leaders in the field are calling on researchers to take more risks and develop new kinds of clinical trial to test different combinations of drugs to treat progressive multiple sclerosis (S10).

A drastic form of treatment that eliminates a patient's immune system and replaces it with fresh stem cells (S11) is offering hope. But many doctors worry that it is too dangerous for widespread use, and it is already attracting unethical and unlicensed imitators.

Until a blockbuster drug treatment is found, patients have to manage their disease. A variety of diets have been developed, aimed at helping to control the symptoms, and they are now finally getting the scientific scrutiny they deserve, despite the difficulties of applying rigorous clinical-trial standards to eating habits (S13). With perseverance, imagination and a bit of luck, progress on treating and curing multiple sclerosis could soon move faster than ever.

This Outlook has been supported by a grant from F. Hoffmann-La Roche Ltd, which has had no control over the editorial content of this activity.

Brian Owens

Contributing editor

CONTENTS

S2 BIOLOGY

A degenerative affliction

The development of multiple sclerosis

S4 AETIOLOGY

Neighbourhood watch

Clusters of multiple sclerosis cases can help reveal the cause

S7 THERAPIES

Progressive steps

Signs of progress in the search for drug treatments

S10 PERSPECTIVE

Who dares, wins

Bibi Bielekova calls for a radical rethink on drug development

S11 STEM CELLS

Stemming the tide

Can a risky treatment help some people who have multiple sclerosis?

S13 DIET

Changing the recipe

It is harder than it looks to test whether dietary changes can help to treat multiple sclerosis

RELATED ARTICLES

S15 Myeloid cells — targets of medication in multiple sclerosis

Manoj K. Mishra & V. Wee Yong

S28 Transcription factor Nr4a1 couples sympathetic and inflammatory cues in CNS-recruited macrophages to limit neuroinflammation

Iftach Shaked et al.

S35 Proton-gated Ca^{2+} -permeable TRP channels damage myelin in conditions mimicking ischaemia

Nicola B. Hamilton, Karolina Kolodziejczyk, Eleni Kougioumtzidou & David Attwell

S40 Interferon- β therapy specifically reduces pathogenic memory B cells in multiple sclerosis patients by inducing a FAS-mediated apoptosis

Fabiana Rizzo et al.

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the Nature Outlook Editorial guidelines available at go.nature.com/e4dwzw

CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol. XXX, No. XXXX Suppl., Sxx–Sxx (2016).

VISIT THE OUTLOOK ONLINE

The *Nature Outlook Multiple Sclerosis* supplement can be found at <http://www.nature.com/nature/outlook/multiple-sclerosis>. It features all newly commissioned content as well as a selection of relevant previously published material.

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES

Site licences (www.nature.com/libraries/site_licences): Americas, institutions@natureny.com; Asia-Pacific, <http://nature.asia/jp-contact>; Australia/New Zealand, nature@macmillan.com.au; Europe/ROW, institutions@nature.com; India, npindia@nature.com. Personal subscriptions: UK/Europe/ROW, subscriptions@nature.com; USA/Canada/Latin America, subscriptions@us.nature.com; Japan, <http://nature.asia/jp-contact>; China, <http://nature.asia/china-subscribe>; Korea, www.natureasia.com/ko-kr/subscribe.

CUSTOMER SERVICES

Feedback@nature.com

Copyright © 2016 Macmillan Publishers Ltd. All rights reserved.

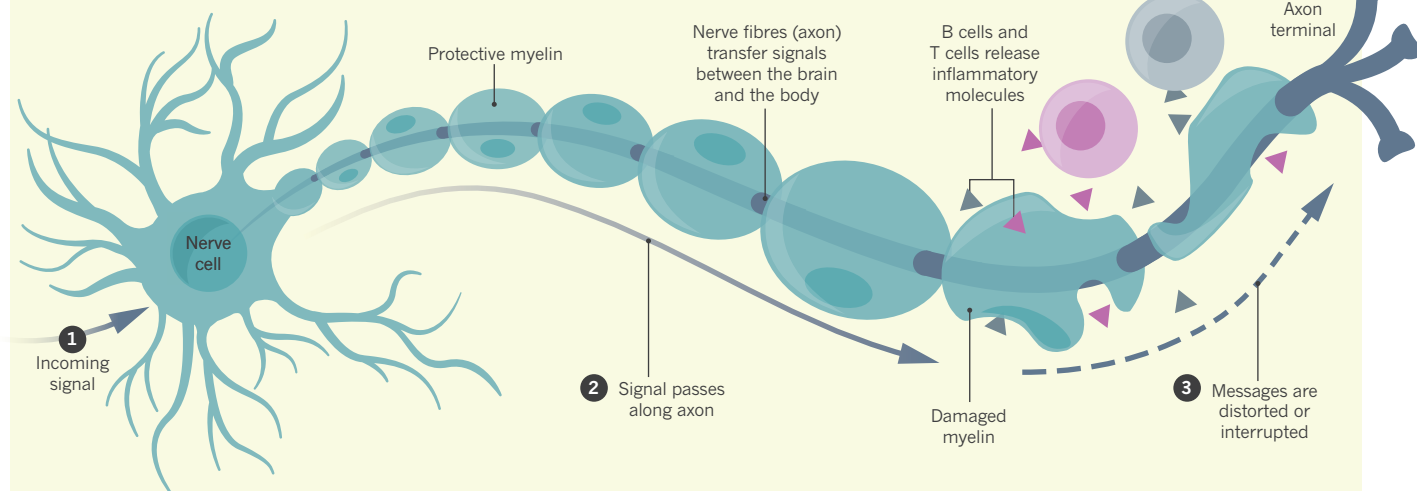
A DEGENERATIVE AFFLICTION

Multiple sclerosis often strikes between the ages of 20 and 40, when people are entering the workforce and raising families.

By Charles Schmidt; illustration by Lucy Reading-Ikkanda.

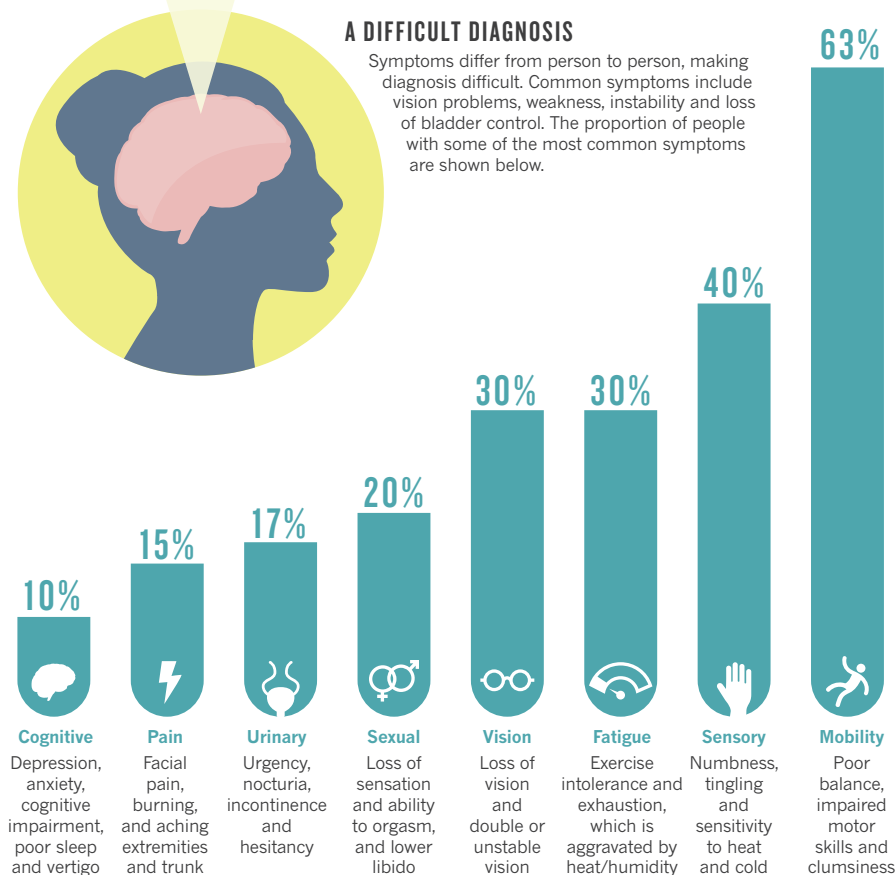
TURNING AGAINST THE SELF

Multiple sclerosis is an incurable illness in which the body's own immune system destroys tissues in the central nervous system. T cells and B cells are thought to remove a protective coating called myelin that wraps around nerve fibres in the brain, spinal column and optic nerve. Exposed fibres are degraded, producing symptoms that vary depending on where the damage occurs.



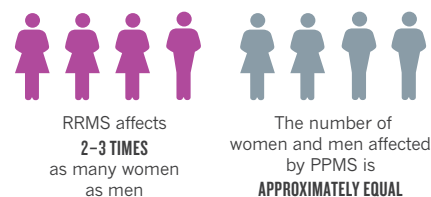
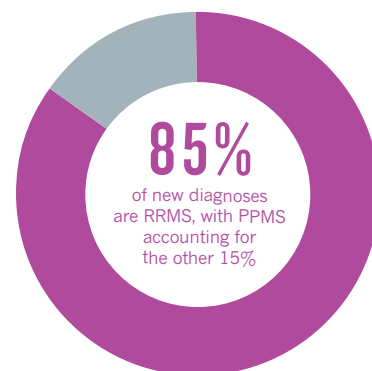
A DIFFICULT DIAGNOSIS

Symptoms differ from person to person, making diagnosis difficult. Common symptoms include vision problems, weakness, instability and loss of bladder control. The proportion of people with some of the most common symptoms are shown below.



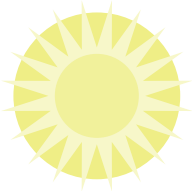
TYPES OF MULTIPLE SCLEROSIS

Relapsing–remitting multiple sclerosis (RRMS) produces attacks followed by periods of remission that can last months or years. It usually gets worse with time. In primary progressive multiple sclerosis (PPMS), symptoms worsen steadily without remission.



WHAT CAUSES MULTIPLE SCLEROSIS?

The cause is unknown, but it is thought to involve an interplay of genetics and the environment. Genetic susceptibility drives the risk, but the contribution of any specific gene seems to be modest.



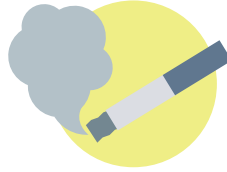
Environmental factors

Living further from the Equator boosts risk, possibly because of reduced exposure to sunlight — the main source of vitamin D.



Epstein-Barr virus

Few people who have been infected with the Epstein-Barr virus develop multiple sclerosis, but nearly everyone with the condition has been infected with the virus.



Smoking

Numerous studies have found higher rates of multiple sclerosis in smokers than in non-smokers.



Genetic factors

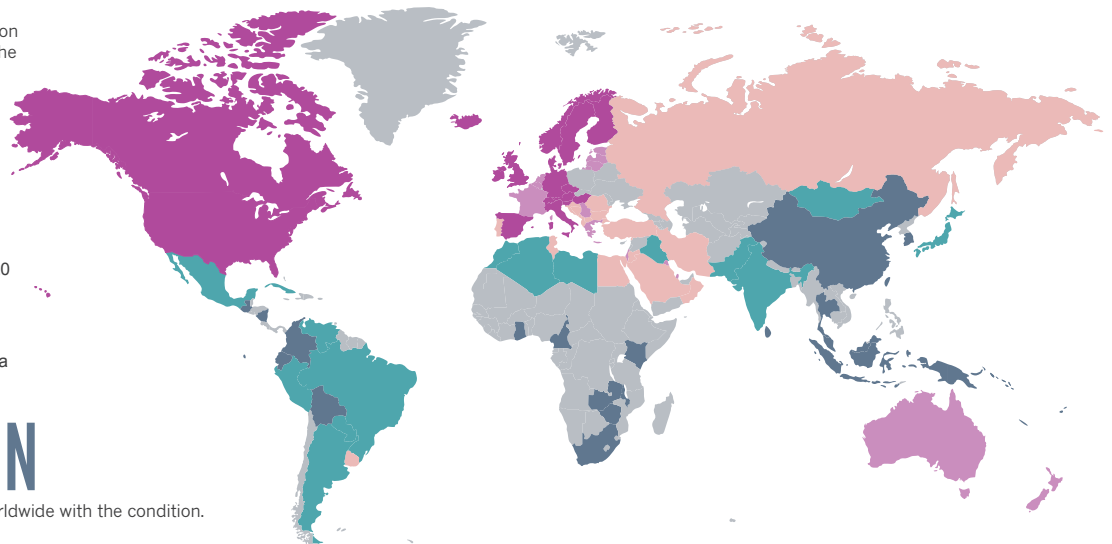
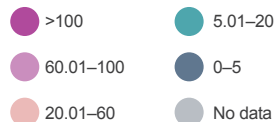
Combined with vitamin D deficiency, the *HLA-DRB1*1501* allele and the *CYP27B1* gene may elevate risk. *IL2RA* gene variants may increase the risk for individuals without a family history of the disease. *IL7R* variants are associated with PPMS.

GLOBAL INCIDENCE

Multiple sclerosis is most common in northern latitudes, reflecting the likely association with reduced exposure to sunlight and vitamin D deficiency.

Prevalence per country

Number of people with multiple sclerosis per 100,000

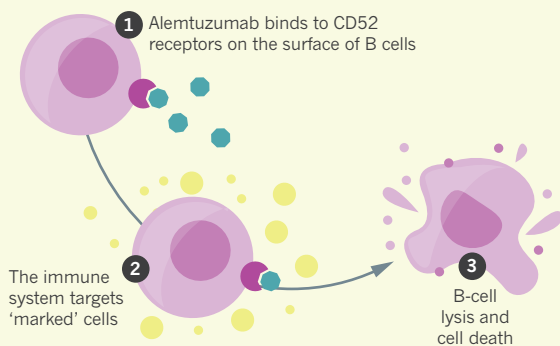


2.3 MILLION

Estimated number of people worldwide with the condition.

DISEASE-MODIFYING THERAPIES

Several drugs have been approved for RRMS, but none for PPMS. Drugs need to be given as soon as possible to slow the pace of neurological damage.

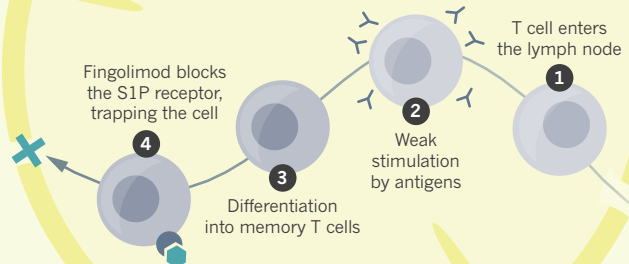


ALEMTUZUMAB

A monoclonal antibody also approved for leukaemia, alemtuzumab binds to CD52, a protein on the surface of mature B cells and T cells, which are then targeted for destruction by the immune system. Widespread B-cell depletion from alemtuzumab treatment may promote the growth of T cells that are less autoreactive and do not cause inflammation.

FINGOLIMOD

T cells in the lymph nodes develop into central memory T cells, which predominate in the cerebrospinal fluid of people with multiple sclerosis. Fingolimod binds to the S1P receptor on T cells, trapping them in the lymph nodes so they cannot reach neurological tissues.



A man with short dark hair, wearing a black jacket over a white t-shirt and grey sweatpants, is sitting in a black motorized wheelchair on a wooden deck. The deck is covered with many fallen yellow and orange autumn leaves. In the background, there is a wooden fence and a house with large windows. The overall atmosphere is quiet and somewhat somber.

Neighbourhood watch

BY CAROLYN BROWN

Emerging evidence points to a viral infection, low levels of vitamin D and genetics as culprits in multiple sclerosis, but how they combine to cause the disease is unclear.

When Jacques Dutrisac arrived at the Ottawa General Hospital in Canada for his first appointment after being diagnosed with multiple sclerosis (MS) in 1991, he met a former high-school classmate in the waiting room. She told him that a third schoolmate had also been diagnosed with the disease. They have since identified 15 people with MS from the neighbourhood where Dutrisac grew up in the late 1960s and 1970s, 14 of whom lived within a kilometre of each other. Four are from the same street, and “there were not even a hundred people on our little street”, explains Dutrisac. Five attended the same elementary school, and seven the same high school. Their

ages span just eight years. In an area that is predominantly English-speaking, ten are French Canadian.

This Ottawa cluster of MS cases has not been analysed statistically to determine whether it is the result of chance alone, but similar MS clusters in the Canadian province of Manitoba have. A study of new cases diagnosed over a 16 year period found several hotspots, mainly in the city of Winnipeg¹. Rates of MS were more than twice as high in the city as they were in the rest of the province — more than three times as high in some neighbourhoods. Clearly something is causing these clusters.

Canada has the highest rate of MS in the world, with 291 cases per 100,000 people². But

rates vary widely from one region of Canada to another^{3–5}. Several studies have concluded that something in the environment in Canada is contributing to MS — indeed, rates of MS in immigrants to Canada quickly rise to approach Canadian levels^{6,7}.

Differences between countries can suggest possible causes. Clusters in particular neighbourhoods, such as those in Ottawa and Winnipeg, raise suspicion about local environmental factors, including outbreaks of infectious disease or genetic differences in multicultural cities, says Mahmoud Torabi, a statistician at the University of Manitoba and an author of the Winnipeg study.

Until recently, researchers could not

TONY FOUHSE

Jacques Dutrisac is one of a cluster of 15 people with multiple sclerosis who grew up in the same neighbourhood in Ottawa, Canada.

University's Montreal Neurological Institute and Hospital. But how they interact to result in MS is not yet clear.

A LIFELONG INFECTION

Inspired by the geographical variation in MS, hundreds of epidemiological studies have been done to identify environmental factors. A wide range of potential causes have been the subject of meta-analysis studies, which pool data from several studies and can include thousands of patients, explains Ioanna Tzoulaki, an epidemiologist at Imperial College London. These meta-analyses have investigated infections, vaccinations, other diseases, surgery, accidents and exposure to toxins. Tzoulaki's team has conducted a review of these meta-analyses⁹. "They looked at every trend or player in these environmental factors," says Røsjo, who was not involved in the review, which he calls the "pinnacle" study of these factors.

The strongest evidence came from studies of links between MS and two signs of EBV infection: a history of mononucleosis, and a biological marker of EBV infection called EBNA1.

This chimes with Dutrisac's experiences. In his teens, he often felt tired, despite being an active adolescent who loved to play ice hockey in the winter. "I would just get tired for no reason sometimes," he says. Fatigue is common in MS, but it is also a symptom of mononucleosis, which strikes mainly adolescents and young adults, with symptoms that also include fever, sore throat, swollen lymph nodes and inflammation of the spleen. Dutrisac was not diagnosed with mononucleosis, but he may have had an EBV infection; many people who are exposed to the virus, especially as children, have no symptoms, and many mild cases go undiagnosed. So although only a minority of people have been diagnosed with mononucleosis, an estimated 90% of people globally have been exposed to EBV by the time they are 40 years old. And the virus is persistent. As Røsjo says, "You don't get rid of it. Once you've got it, you have it for life."

There are three broad theories to explain how EBV infection could result in MS, explains Bar-Or. One possibility is that the EBV infects the central nervous system (CNS), especially the brain. The virus may then kill oligodendrocytes, the cells that produce the myelin sheath — the insulating envelope that coats the nerve fibres (axons). Myelin speeds up the nerve impulses travelling through the axons and is essential for normal nerve function.

In the CNS, the infection may also trigger an immune response in which a type of immune cell known as CD8 tries to kill the virus but also damages oligodendrocytes and neurons. Bar-Or admits that this theory is controversial, because although some researchers have found the virus in the CNS, others have been unable to replicate the finding.

A second theory involves molecular mimicry, in which the immune system, Bar-Or

says, may mistake myelin basic protein (MBP), which is important for the myelination process, for a similar piece of the invading EBV. The immune system's T cells have receptors that can recognize both the virus and MBP, and when the T cells are activated during the immune response to the virus, they could attack MBP as well as EBV in parts of the brain, resulting in a loss of myelin.

A third possible mechanism involves immune cells known as B cells. In normal

“The body of evidence is difficult to get a grip on, it’s so huge.”

immune responses, there are interactions between B cells, myeloid cells and T cells when responding to invading bacteria and viruses, and reversing the immune response when it is no longer needed. Any imbalance in this delicate dance can cause a strong or long-term response that goes beyond its original purpose and injures the body by interfering with basic cell functions, says Bar-Or. Perhaps EBV causes such an imbalance in MS by activating the B cells that they no longer perform their crucial regulatory functions, but instead start to promote inflammation. There is some support for this theory because MS treatments that remove B cells have succeeded in decreasing MS relapses, says Bar-Or.

VITAMIN D LINK COMES TO LIGHT

World maps of MS (see page S2) show that its prevalence increases as you move further from the Equator. This geographical distribution suggests that MS may be linked to the lower levels of vitamin D found in people living in regions with long, dark winters. But evidence supporting this theory has come mainly from observational studies, which cannot show cause and effect, says Brent Richards, an endocrinologist who studies vitamin D and genetics at McGill University in Montreal. For example, it is unclear which way the causality runs — rather than low levels of vitamin D causing MS, it is possible that the disease somehow lowers vitamin D levels, Richards says. Another complication is that “vitamin D is associated with a host of healthy lifestyle behaviours,” he says. “If you go outside, you tend to be healthier and your vitamin D levels tend to be higher.” In this

pinpoint why Canada in general, and specific places in particular, are so rife with this neurodegenerative disease^{6–8}. But an international research effort — including database analyses, epidemiology, microbiology, genetics and immunology — is now yielding a growing body of evidence on the causes of MS. “It’s difficult to get a grip on, it’s so huge,” says Egil Røsjo, a neurologist at Akershus University Hospital near Oslo, who is also completing a doctorate on MS aetiology. “There are lines of evidence that are pointing more or less in the same direction from all these different angles.”

And they are pointing to three main culprits: Epstein–Barr virus (EBV), which is a herpes-virus that causes mononucleosis (glandular fever); low levels of vitamin D; and genetic variants that increase susceptibility to MS.

But each of these potential causes is common, and MS is quite rare, so none of these alone is likely to hold the key. “The development of MS is based on combinations of these,” argues Amit Bar-Or, a neurologist at McGill

case the healthy behaviours may prevent MS, and the geographical variation would need another explanation.

Demonstrating a causal link through a randomized controlled trial (RCT) of vitamin D would be complex, lengthy and costly, and such a trial may not be funded anyway because it would not test a drug. Richards' team tackled this problem by conducting a study that he calls "nature's RCT". They analysed large databases to find genetic variants that lowered vitamin D levels and then, by using another genetic database of people with MS, determined the MS risk in people with these variants. They found a strong risk of MS in people who have naturally low levels of vitamin D — and the risk increased as the vitamin D level decreased¹⁰. "The person had their genetic make-up before they had MS. That's important, because the disease cannot influence the genes," he says.

The mounting evidence of the link with vitamin D has sparked discussion about how low levels could lead to MS. Bar-Or points out that vitamin D has been shown to affect the immune response in many of the cell types involved in MS, including myeloid, B and T cells. "There's an enormous amount of literature that associates vitamin D deficiency with risk of disease," adds Richards.

Røsjo thinks that this points to the connection with EBV. His research follows up a theory¹¹ that vitamin D may boost immune responses to EBV and, conversely, that a lack of vitamin D may lead to poor responses to EBV. He thought that a weak immune response would increase levels of the biomarker EBNA1. "This in some way reflects increased risk of MS disease development," says Røsjo, although the molecular basis is still unknown. But *in vitro* studies have shown that the form of vitamin D that is active in the body, 1,25-dihydroxyvitamin D₃, inhibits the production of antibodies from B cells, which are the cells infected by EBV.

Røsjo's study has borne out this theory. In a long-term placebo-controlled study of high-dose vitamin D supplementation in people with MS, Røsjo found¹² that levels of antibodies against EBNA1 fell significantly over the first 48 weeks of the study, but then rebounded, although never to their original levels, over the following 48 weeks. Røsjo says that this study was the first indication of an interaction between vitamin D and EBV. "There might be a short-term effect" of vitamin D on MS, he adds, "but it's lost over time."

The study also looked at two other viruses in the same herpesvirus family as EBV — varicella zoster virus, which causes chickenpox and shingles, and cytomegalovirus — and found that levels of antibodies to these viruses were unaffected by vitamin D. "There is something special, something unique with EBV that we cannot explain," Røsjo concludes.

Røsjo found that even large doses of

vitamin D do not help patients with established MS in the long term, but he thinks that it may have a role during initial EBV infection. "Vitamin D and EBV during adolescence might be some of the most important things when it comes to deciding MS risk," he says.

FAMILY FORTUNES

One of the big puzzles about MS is that it runs in some families, but it can also pop up sporadically in people with no family history of the disease. Meagan McEwen, who campaigns for people with MS in Ottawa and has the disease, and her boyfriend, who also has it, embody this puzzle. "In my boyfriend's family, there are many other people who have MS. That's not true in my family," she says. This paradox has fuelled hundreds of studies worldwide.

One such study focused on the risk factors associated with the onset of MS in Canadian

"The genetic contribution doesn't explain even half the risk."

children. A team of researchers, including Bar-Or, investigated several major risk factors, including EBV, vitamin D and a specific variant of the *HLA-DRB1* gene, which is one of the human leukocyte antigen (HLA) complex of genes that are involved in the immune response¹³. "One of the benefits of studying children is that there's a relatively short time between the time they are born and the time they develop disease," explains Ruth Ann Marrie, a neurologist at the University of Manitoba in Winnipeg and one of the co-authors. "There is less time for either extraneous or relevant events to occur."

Bar-Or and Marrie found that the genetic variant most strongly associated with MS around the world, *HLA-DRB1*15*, doubled the risk of children having MS. Since that study was published in 2011, the number of genetic variants known to affect MS risk has climbed from roughly 50 to about 200.

Bar-Or stresses that none of these MS-associated variants is abnormal — they are not 'disease genes' like the mutations that cause cystic fibrosis, for example. "There are many, many more healthy people who have one or more combination of those very same

variants than there are people who develop MS," he says. It seems that environmental factors or epigenetics — changes to gene expression that do not alter the genetic code — may explain the different outcomes, even for people with the same genetic make-up. "Two individuals who may have the identical template end up manifesting their genes in different ways based on epigenetics," says Bar-Or.

Genetic susceptibility can also vary geographically, Marrie points out, which may help to explain some clusters. A study of genetic risk factors for MS in Ireland, for example, showed that the percentage of individuals with the genetic variants that put them at risk of MS was substantially lower in one county than in another one nearby¹⁴. "The genetic contribution doesn't explain even half the risk," cautions Bar-Or, so it may be possible to prevent the disease by targeting the environmental factors involved.

Marrie thinks that the future direction of MS research will be into how these factors interact. "Now that we have identified multiple potential factors, we will work on the biology of the relationship to understand how these factors work together, and whether order of exposure is important," says Marrie. But she and others admit that definitively proving the causal role of any of these factors may be impractical or even impossible.

None of the researchers are ready to say what factors might be causing the Canadian clusters. Could it be the long, dark winters? A genetic variant common in French Canadians? A local EBV outbreak? Some combination of all three?

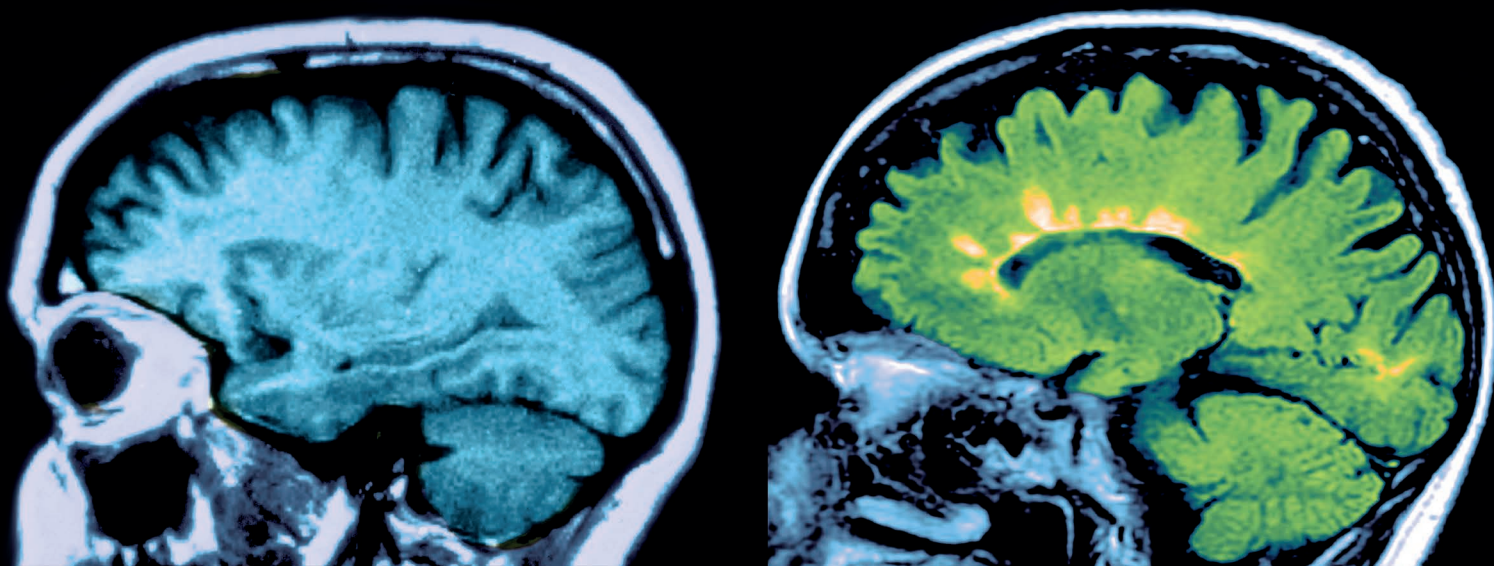
Dutrisac hopes that information about his cluster can help researchers to find the answer. "There's something out there. We need to find the cause," he says. ■

Carolyn Brown is a freelance writer based in Ottawa, Canada.

1. Torabi, M., Green, C., Yu, N. & Marrie, R. A. *Neuroepidemiology* **43**, 38–48 (2014).
2. Multiple Sclerosis International Federation *Atlas of MS* (MSIF, 2013).
3. Poppe, A. Y., Wolfson, C. & Zhu, B. *Can. J. Neurol. Sci.* **35**, 593–601 (2008).
4. Beck, C. A., Metz, L. M., Svenson, L. W. & Patten, S. B. *Mult. Scler.* **11**, 516–519 (2005).
5. Warren, S. & Warren, K. G. *Neurology* **43**, 1760–1763 (1993).
6. Guimond, C. *et al. Mult. Scler.* **20**, 1182–1188 (2014).
7. Orton, S.-M. *et al. J. Neurol. Neurosurg. Psychiatry* **81**, 31–36 (2010).
8. Orton, S.-M. *et al. Lancet Neurol.* **5**, 932–936 (2006).
9. Belbasis, L., Bellou, V., Evangelou, E., Ioannidis, J. P. A. & Tzoulaki, I. *Lancet Neurol.* **14**, 263–273 (2015).
10. Mokry, L. E. *et al. PLoS Med.* **12**, e1001866 (2015).
11. Holmoy, T. *Med. Hypoth.* **70**, 66–69 (2008).
12. Røsjo, E. *et al. Mult. Scler. J.* <http://dx.doi.org/10.1177/1352458516654310> (2016).
13. Banwell, B. *et al. Lancet Neurol.* **10**, 436–445 (2011).
14. McGuigan, C. *et al. J. Neurol.* **252**, 1245–1248 (2005).

Progressive steps

New drugs are beginning to show promise for people with one of the less common, and harder to treat, forms of multiple sclerosis.



BY ELIE DOLGIN

Mitch Sturgeon has tried more than half a dozen therapies to treat his multiple sclerosis (MS), from immune modulators to chemotherapy drugs and vein-widening surgical procedures. None of them seem to have slowed the disease's advance. He did experience a temporary reprieve about a decade ago, thanks to a monoclonal antibody called rituximab, which he received during a clinical trial. But when the two-year study ended in 2007, with the conclusion that the therapy was no better than a placebo¹, it was back to the drawing board for Sturgeon, a former chemical engineer from South Portland, Maine.

The reason that nothing seems to work for Sturgeon is that he has primary progressive MS, a form of the disease marked by steadily worsening neurological function from the onset of symptoms. There are currently 14 drugs approved by the US Food and Drug Administration (FDA) to treat MS, including six that hit the market in the past five years. Yet all these agents combat the clinical attacks associated with the more common relapsing-remitting MS. No medication has ever been shown to help the 10–15% of patients with MS, like Sturgeon, whose physical decline is continuous and unrelenting. And only one drug — a 16-year-old immunosuppressant that is rarely prescribed because of its harsh side effects — has been approved for the secondary

progressive form of MS that eventually develops in most patients who first experience the relapsing type.

That dearth of therapies could be about to change. Before the end of December, the FDA is expected to decide on the fate of a new drug called ocrelizumab, which works in a similar way to rituximab by depleting the B cells in the immune system that are crucial for promoting inflammation in MS. The drug is also under review in Europe, Australia and elsewhere. In phase III trials, ocrelizumab proved superior to an existing drug that is commonly used to reduce the frequency of relapses for people who experience them. And promisingly for patients such as Sturgeon, it outperformed a placebo in preserving physical function in primary progressive disease.

If ocrelizumab is approved, Sturgeon expects to try it. For almost two years he has been taking high doses of a vitamin called biotin, even though it has no effect on his disease. "I haven't gotten off it because I have nothing else lined up right now," Sturgeon says. Ocrelizumab could fill that gap — and other drugs might soon be on the way. "The door is ajar for the first time," says Stephen Hauser, a neurologist at the University of California, San Francisco. "But this is only the beginning of effective therapies for progressive MS."

MS is usually thought of as an immune-mediated disease in which misdirected T cells, B cells and other inflammatory drivers infiltrate the brain and destroy the protective

myelin coating around nerve fibres, disrupting signalling to and from the brain and spinal cord. But there is also a neurodegenerative component, and this is particularly pronounced in progressive forms of MS, which cause widespread destruction in the brain's white and grey matter.

A scan of a healthy brain (left) and one with multiple sclerosis, showing demyelinating lesions (yellow).

LEARNING FROM FAILURE

The differences between progressive and relapsing disease remain a matter of debate, but most researchers agree that relapsing MS is characterized by episodes of intense inflammation and demyelination, whereas in progressive MS the inflammation is lower grade and chronic, with neurons dying a slow death from multiple, currently unknown, causes.

Another distinction lies in the integrity of the blood–brain barrier. In relapsing disease, this boundary becomes permeable, allowing immune cells — and the drugs that target them — to pass from the bloodstream into the brain. By contrast, in progressive MS, the blood–brain barrier remains intact, confining the inflammation to the brain, which fuels atrophy and neuronal loss. This firm barrier, and the resulting inability to reach the inflammation with medicines, could explain, at least in part, why nearly every trial of an immunomodulating therapy that works for relapsing MS has failed for patients with progressive

disease — the drugs simply do not reach their target in the brain. An alternative theory is that progressive MS does not result from inflammation at all, but rather from the degeneration that follows the early inflammatory injury.

There is another problem, however. Trials for progressive MS have tended to include everyone with the disease, regardless of whether they show signs of the active, infiltrating type of inflammation that is known to be affected by immune-targeted therapies. This type of inflammatory lesion, which can be detected on a magnetic resonance imaging (MRI) brain scan, is common in relapsing disease but is seen in only a small subset of patients with progressive MS — generally younger people with more recently diagnosed disease. For this reason, drug companies with immune-modulating agents have begun to redesign their studies of progressive MS to enrol more of the patients who are most likely to benefit. This strategy has yielded more success in clinical trials, but it means that these therapies might not be appropriate for everyone with progressive MS.

Ocrelizumab is a case in point. Like rituximab, ocrelizumab targets a protein called CD20 on the surface of B cells. What sets the new drug apart is that it can eliminate B cells with greater efficiency. What's more, whereas rituximab includes mouse components, ocrelizumab has been modified to more closely resemble proteins produced naturally by the human body, making it less likely to trigger a drug-related immune reaction.

In the study population as a whole, rituximab did not outperform a placebo¹. But Roche, the Swiss pharmaceutical company behind the drug, noticed some improvements in younger patients with primary progressive MS, particularly in those with active inflammation. Taking those lessons on board, the company adjusted the enrolment criteria and other design features for its late-stage ocrelizumab trial to study younger patients who were more likely to have active disease — and the change paid off.

In 2015, Roche revealed the results of its

732 person trial, which showed that regular infusions of ocrelizumab reduced the risk of disease progression by 24% compared with a placebo. It significantly slowed declines in walking ability, upper-limb function and brain volume for study participants, with no major safety concerns.

"This is the first positive phase III treatment trial in primary progressive MS," says Peter Chin, group medical director for neuroscience at Genentech, a US subsidiary of Roche. "To show consistency of effect across a number of different disease-progression measurements is a very meaningful result."

But Hauser, who was a trial investigator for the drug, warns that the benefit was "really modest" and that ocrelizumab might not work for older patients in the more advanced stages of disease. "We have to be careful how this is communicated to patients," he says.

Novartis, another Swiss pharmaceutical company, found a similar stratification

"We need to move beyond drugs that work on T cells and B cells."

of benefit in its trial of siponimod for secondary progressive MS. Siponimod is an experimental drug that prevents the transport of immune cells to sites of inflammation in the brain. It targets the same protein on the surface of white blood cells as another Novartis drug called fingolimod, which has been approved for relapsing MS under the brand name Gilenya. In a 970 person study published earlier this year², led by Fred Lublin, director of the MS clinic at the Icahn School of Medicine at Mount Sinai in New York City, fingolimod showed no benefit for primary progressive MS. But a 1,651 person trial of siponimod to treat secondary progressive MS found that it reduced the risk of three-month disability progression by 21% compared with a placebo.

The overall effect is small, admits lead trial investigator Ludwig Kappos, a neurologist at University Hospital Basel in Switzerland. But "it was reassuring and a positive signal that we

also saw a benefit in people who were quite advanced in the progressive phase of MS and already had severe disability," he says. Kappos presented the findings at the 2016 congress of the European Committee for Treatment and Research in Multiple Sclerosis in September.

REUSING DRUGS

Lublin is buoyed by the positive trial results for ocrelizumab and siponimod, both of which he helped to test. "What's really exciting is we've had initial success for two very difficult forms of MS to treat," Lublin says. However, apart from those two immune-modulatory agents, there is not much in the late-stage pipeline for progressive MS.

But there are therapies in earlier stages of clinical testing that are focused on tackling the brain damage that is the hallmark of progressive MS (see "Therapies in the pipeline"). Some involve stem cells that are designed to restore lost nerve function (see page S11). Most of the investigations focus on neuroprotective agents that have already been approved to treat other diseases and that might be beneficial in progressive MS as well — or at least point to molecular pathways that could be targeted.

The French biotechnology company AB Sciences, for example, is testing whether a cancer drug called masitinib, which is used mainly in dogs, can help patients with primary and secondary progressive forms of MS. And a group led by Jeremy Chataway at University College London is evaluating whether any of three repurposed human medicines — a heart-disease drug, an antidepressant and a treatment for amyotrophic lateral sclerosis (also known as motor neurone disease) — will help with the secondary progressive stage of disease. Chataway explains that the three drugs in his trial were chosen after an exhaustive review³ of all the licensed neuroprotective therapies that can be taken orally and that have some preliminary data on whether they can help in MS. "We looked at thousands and thousands of reports to really drill down to this short list of drugs that could be useful against the pathology of progressive multiple sclerosis," he says.

One of the drugs that was shortlisted by Chataway's group, but not included in his trial, was ibudilast, which is marketed in parts of Asia to treat asthma and stroke-related symptoms. Ibudilast has previously shown no benefit in people with relapsing MS, but there was some indication that it might protect the nervous system from damage because of its ability to block an enzyme involved in brain inflammation⁴. That observation prompted Robert Fox, medical director of the Cleveland Clinic's MS centre in Ohio, to launch a trial to test ibudilast for primary progressive MS. "This is one of the few molecules that has empirical evidence from MS itself that it slows atrophy progression," Fox says.

The 250 person ibudilast study is exploring the drug's effects on brain shrinkage, which is

THERAPIES IN THE PIPELINE

Several drugs are currently being developed to treat progressive multiple sclerosis. Some are repurposed drugs that are already being used to treat other diseases, and many have a neuroprotective effect.

Drug	Mechanism of action	Disease course under investigation	Phase
Ocrelizumab	CD20 antagonist	PPMS	Regulatory filing
Siponimod	Sphingosine-1-phosphate receptor modulator	SPMS	III
Masitinib	Tyrosine kinase inhibitor	PPMS, SPMS	III
High-dose biotin	Vitamin B7	PPMS, SPMS	III
Ibudilast	Phosphodiesterase inhibitor	PPMS, SPMS	II
Idebenone	Coenzyme Q10 analogue	PPMS	II
Amiloride	Potassium-sparing diuretic	SPMS	II
Riluzole	Glutamate-release inhibitor	SPMS	II
Fluoxetine	Selective serotonin reuptake inhibitor	SPMS	II

PPMS, primary progressive multiple sclerosis; SPMS, secondary progressive multiple sclerosis.



Fred Lublin (left) is pleased with the progress being made to treat progressive multiple sclerosis.

the standard approach in mid-stage clinical trials of MS therapies that are seeking proof of principle that the drug is working in the brain. But unlike most trials of its kind, this study also includes several advanced imaging techniques and tests of disability so that, even if ibudilast shows no benefit, the investigators will learn something about their outcome metrics through direct head-to-head comparisons.

Another idea for improving outcome measures comes from Bibi Bielekova, head of the neuroimmunological diseases unit at the US National Institute of Neurological Disorders and Stroke in Bethesda, Maryland. Bielekova's team statistically compared 58 outcome measures in a group of 98 untreated people with primary progressive MS, and concluded that most brain-scanning metrics are unreliable surrogates for symptom progression (see page S10). The researchers integrated four different clinical scales into a single disability score called CombiWISE, and it outperformed other measures for detecting clinical deterioration⁷.

"I have zero doubt that CombiWISE is the best thing we currently have for screening treatments," says Bielekova, who is currently using the metric in an 85 person trial of the antioxidant idebenone for primary progressive MS.

BACK TO BASICS

Most researchers who study progressive MS remain cautiously optimistic that at least some of the experimental therapies for this form of the disease will prove their worth. But the challenges are daunting, largely because so much remains unknown about the basic pathology of progressive MS. "We still need to

better understand the mechanisms of progression," says Tim Coetzee, chief advocacy, services and research officer at the US National Multiple Sclerosis Society in New York City.

To address these knowledge gaps and accelerate progress in new therapeutic directions, the society teamed up in 2013 with other research and advocacy non-profit groups from around the world to launch the Progressive MS Alliance. This €22.4 million (US\$24.0 million) initiative will be "a catalyst that brings together the best institutions worldwide," says Alan Thompson, dean of brain sciences at University College London and chair of the alliance's scientific steering committee. "This has never happened in the world of MS before," he says.

In 2014, the alliance funded a series of 20 small-scale research projects on topics ranging from genetics and disease models to proof-of-concept trials. And earlier this year it handed out three larger grants worth a total of €12.6 million for drug-discovery efforts and biomarker development. Francisco Quintana, an immunologist at the Brigham and Women's Hospital in Boston, Massachusetts, has received both kinds of award.

Quintana used his €75,000 grant from the Progressive MS Alliance to demonstrate in a mouse model of MS that miglustat, a drug currently used to treat Gaucher's disease, can stop the activation of astrocytes, a type of immune cell that promotes inflammation in the brain. Now, thanks to a four year, €4.2 million award, Quintana is trying to find more drugs that can cross the blood-brain barrier and block astrocyte function or disrupt a metabolic pathway

involved in the activation of astrocytes.

"We need to move beyond drugs that work on T cells and B cells," Quintana says. Modulating those cells may help to treat relapsing MS, but progressive disease is driven by innate immune cells such as microglia and astrocytes, and it's "those cells we should be targeting to arrest chronic progression," he says.

Research efforts such as Quintana's could eventually help the million plus people worldwide who live with the disability of progressive MS. But as the years pass and his condition continues to deteriorate, Sturgeon, who now relies on a powered wheelchair and voice-controlled devices, is losing hope of researchers finding a therapy that can help someone in his advanced disease state. He plans to take ocrelizumab next year after its anticipated approval, but he realises that the likelihood of it working for him is "pretty low".

"I'm walking a fine line between hope and acceptance," says Sturgeon, who chronicles his experiences in a blog called *Enjoying the Ride*. But there is one reason to maintain hope. "We don't seem to be the forgotten group anymore," he says. "There's some attention coming our way — and that helps." ■

Elie Dolgin is a science writer based in Somerville, Massachusetts.

1. Hawker, K. et al. *Ann. Neurol.* **66**, 460–471 (2009).
2. Lublin, F. et al. *Lancet* **387**, 1075–1084 (2016).
3. Vesterinen, H. M. et al. *PLoS ONE* **10**, e0117705 (2015).
4. Barkhof, F. et al. *Neurology* **74**, 1033–1040 (2010).
5. Kosa, P. et al. *Front. Neurol.* **7**, 131 (2016).

PERSPECTIVE



Who dares, wins

It is time for a bolder approach to developing drugs for progressive multiple sclerosis, says **Bibi Bielekova**.

The economic and human cost of central nervous system (CNS) disorders is enormous¹, and neurodegenerative diseases such as progressive multiple sclerosis (MS) represent the fastest-growing portion of this cost. But drug development for these diseases has so far been slow, expensive and mostly unsuccessful². It is time for a rethink.

Current disease-modifying treatments for MS, all of which target the immune system, have convincingly demonstrated that processes other than the formation of new lesions underlie much of the disability that develops as MS progresses. Many potentially pathogenic cellular processes have been implicated, but we still do not know which, if any, are actually involved.

It is unlikely that such a complex disease can be stopped by targeting a single process, so treating progressive MS will probably require a combination of carefully targeted therapies. But making such personalized combination treatments a reality will require a major overhaul of approaches to drug development and clinical trials, including new ways of measuring their success.

Evidence that this is feasible comes from cardiovascular diseases, where this kind of individualized 'polypharmacy' is already a mainstream approach. Using biomarkers to identify multiple pathogenic processes, such as hypertension, hypercoagulability or hyperlipidaemia, allows physicians to provide the optimal combination of drugs.

Compare that with MS care, where someone with an MS diagnosis is offered one of many available treatments based on side-effect profile and perceived efficacy, but not on the mechanisms responsible for that person's disease. When follow-up examinations reveal the development of a new CNS lesion, the old medicine is withdrawn and a new one is provided, equally blindly. Obviously, such a strategy is suboptimal: if there are multiple pathogenic processes, they need to be targeted simultaneously by several drugs.

Unfortunately, this need is at odds with our current mode of drug development, which strives to demonstrate the clinical efficacy of each candidate drug individually. If multiple pathogenic processes contribute to the destruction of CNS tissue, then targeting any single one of them, even with a highly effective therapy, will have only a small effect on the clinical outcome that is impossible to measure reliably in small, proof-of-principle phase II trials.

Despite our lack of knowledge about the molecular mechanisms that underlie MS progression, we continue to use only insensitive clinical or imaging outcomes that require the study of hundreds of patients for at least two years to screen one drug at a time. This expensive and slow strategy has had some partial successes: systemically administered immunomodulatory agents have shown some efficacy when treating progressive MS. When given early enough, generally before the age of 50, such treatments may reduce the accumulation of disability by up to 25%. But how do we target the remaining 75%?

Whenever a research goal seems beyond our reach, it is time to examine whether the path we have taken assures the highest probability of success. At some point, the prohibitive cost of screening MS drugs in de facto phase III trials will force us to adopt other, more creative solutions.

Sampling cerebrospinal fluid (CSF) could provide comprehensive molecular information about processes in the CNS, but current clinical practice avoids it because lumbar puncture is considered to be an invasive procedure. Some argue that adding lumbar puncture to clinical trials would severely limit patient participation. But in my extensive experience of running MS clinical trials in which lumbar punctures are as frequent as every 3–6 months, fewer than 10% of patients decline to participate because of the procedure. My team has been able to measure with sufficient statistical power the varied pharmacodynamic effects of drugs being tested in as few as 15 patients treated for as

little as 3 months^{3,4}. This greatly outperforms the most sensitive clinical or structural-imaging MS outcomes⁵ and is analogous to using biomarkers such as blood pressure or lipid levels to screen drugs for their potential efficacy in cardiovascular diseases.

The major drawback of using CSF biomarkers for drug development is not the invasiveness of a lumbar puncture, but the uncertainty about the predictive power of the biomarkers themselves. This gap in the knowledge can be filled only through clinical trials in which any early effects on biomarkers can be linked to later clinical outcomes. Adding biomarkers to the early stages of clinical trials can rapidly eliminate any treatments that have limited effects on the targeted process⁴, and can provide solid data for the optimization of doses and the number of participants required for phase III trials of promising agents. And finally, broad measurements of biomarkers in clinical trials would eventually improve the knowledge base sufficiently to allow the assembly of effective drug combinations during the drug-development process and later in clinical practice.

If we are to successfully tackle the burden of neurological diseases before their costs overwhelm us, it is time for neurology to become a more daring clinical discipline. ■

Bibi Bielekova is an investigator in the Neuroimmunological Diseases Unit at the Intramural Research Program of the US National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, which supports her research programme.
e-mail: bibi.bielekova@nih.gov

1. Olesen, J. *et al.* *Eur. J. Neurol.* **19**, 155–162 (2012).
2. Kola, I. & Landis, J. *Nature Rev. Drug Discov.* **3**, 711–716 (2004).
3. Lin, Y. C. *et al.* *Ann. Clin. Transl. Neurol.* **2**, 445–455 (2015).
4. Komori, M. *et al.* *Ann. Clin. Transl. Neurol.* **3**, 166–179 (2016).
5. Kosa, P. *et al.* *Front. Neurol.* **7**, 131 (2016).

The author declares competing interests, see go.nature.com/2ebdmou for details.

IT IS UNLIKELY
THAT SUCH A
COMPLEX
DISEASE CAN
BE STOPPED BY
TARGETING A
SINGLE
PROCESS.



Jennifer Molson had severe multiple sclerosis until a stem-cell therapy triggered a recovery, and she now works at the Ottawa hospital where she was treated.

STEM CELLS

Stemming the tide

Could a high-risk treatment play a part in tackling multiple sclerosis?

BY ASHER MULLARD

When neurologist Mark Freedman and haematologist Harry Atkins started planning a radical multiple sclerosis (MS) experiment in 1999, they set the wheels in motion for a spectacular failure. They learned nothing about how MS develops — but still achieved something remarkable.

The two doctors at the University of Ottawa in Canada wanted to reset a patient's immune system in much the same way that you reboot a computer, hoping to watch the autoimmunity redevelop and thereby pinpoint the elusive origins of MS. The restart button was dangerous: a concoction of potentially fatal toxins that would destroy the aberrant immune cells that drive the neurodegenerative disease, followed by a transplantation of the patient's own haematopoietic stem cells (which give rise to blood cells) to protect against subsequent infection. The researchers hoped that done cautiously, this extreme approach, called autologous haematopoietic stem-cell

transplantation (aHSCT), might also buy their patients a little extra time.

Freedman and Atkins have yet to see how immune cells turn against the central nervous system in MS. “We failed, miserably,” says Freedman, with a smile. Instead the pair had stumbled across something much more useful: a powerful treatment for the disease. Earlier this year they reported that of the 24 patients who have received aHSCT at the Ottawa Hospital, 23 have not relapsed¹.

“We took patients who were really going downhill fast, and we stopped the disease dead in its tracks,” says Freedman. The responses are durable too, having lasted more than 14 years in some patients. Most of the patients stopped getting sicker, but half a dozen have made remarkable neurological recoveries.

Other hospitals have reported similar success using aHSCT to treat MS. A recent meta-analysis of 15 studies since 1995, involving 764 patients, found that 67% of aHSCT recipients had no evidence of disease activity 5 years after treatment².

Even so, many neurologists have reservations about the high-risk treatment, which haematologists initially developed as a last resort to treat blood cancers. And trials of aHSCT in MS have all been small and uncontrolled, prompting some neurologists to argue that it is premature to embrace the results.

“We’ve had many treatments that have appeared promising when used in uncontrolled trials, but that have failed in larger trials,” warns Christopher Bever, a neurologist at the University of Maryland Medical Center in Baltimore. Already, he says, exploitative, unlicensed clinics around the world are profiteering from the enthusiasm for stem cells.

RISKY BUSINESS

There is no disputing that the therapy carries great risk. The drugs that knock out the immune system bring patients to the brink of death — and around 2% of aHSCT-treated patients die from transplant-related causes². Although this death rate is falling, reflecting a better understanding of how to deploy the

treatment in the clinic, recovery remains slow and painful, and the side effects can be lifelong. “It’s not for everyone,” says Freedman.

One person who did benefit, though, is Jennifer Molson. She was diagnosed with MS in 1996, after pins and needles in her left hand made her drop a jug of milk on the floor. Five years later, at the age of 26, her disease was so aggressive that she needed help bathing, getting dressed and eating. She had quit her job, given up her dreams of becoming a police officer, and was confined to a wheelchair.

A grim magnetic resonance imaging (MRI) scan eventually provided a silver lining. “It looked like someone had taken a cheese grater to my brain,” she says. Freedman, who was one of her doctors, turned to her and said “Congratulations, you are sick enough now. Now we can try and do something for you,” she recalls.

In June 2002, Molson became the fifth MS patient to receive aHSCt at the Ottawa Hospital. Freedman and his colleagues first harvested stem cells from her bone marrow that they would later use to bolster her immune system. For 10 days, they blasted her with chemotherapy to kill off the faulty immune cells that were responsible for her decline; the rest of her immune system was collateral damage. On day 11, the doctors reinjected her with the stem cells to stave off what could otherwise be deadly infections.

She hit rock bottom during the subsequent month in the hospital, and spent more than a year at home recuperating. “I had no idea how sick I was going to be,” she says. She still had the neurodegenerative losses of MS, but she also started vomiting on a daily basis and was plagued by ‘chemo fog’ — thinking and memory problems brought on by the chemotherapy drugs. Her energy was gone, as was her hair. She came down with a painful case of shingles and a blood infection. “I went into a bit of depression,” she says. “What did I do to myself? I’d wonder.”

But Molson was one of the lucky ones. The next patient to enter the trial died 62 days after transplantation from treatment-related sepsis and liver injury.

Eventually, Molson made a remarkable recovery. When she started the trial, she could not stand up on her own. Last summer, she went paddle boarding on the choppy, chilly Ottawa river. She still has to take drugs for heartburn and for chemotherapy-induced menopause, but she has been free from MS relapse for 14 years. “I like to use the word ‘cure’,” says Molson, who now works as a research assistant in the hospital where she was treated.

In total, 35% of the patients in Freedman’s

trial have enjoyed sustained improvements on an expanded disability status scale, which quantifies MS disability.

No one really knows what is driving these recoveries, although Freedman suspects that remyelination might be occurring. In MS, inflammatory flare-ups strip neurons of their protective myelin sheaths, leaving them vulnerable to degeneration. Cells called oligodendrocytes can remyelinate damaged neurons, but this regenerative ability tends to fade as the disease progresses. Researchers have struggled, and so far failed, to use drugs to induce remyelination³.

The stem cells used in aHSCt are intended to protect patients from infection, and are not thought to be involved in remyelination. But preliminary imaging data nevertheless suggest that the remyelination pathway might be activated in some aHSCt recipients.

DIVIDED OPINION

Neurologists remain divided over how to roll out aHSCt to patients with MS, however. “The bottom line for me is that aHSCt is unproven,” says Bever. He calls the results “striking”, but argues that they need to be handled with care until they are replicated in a larger, controlled clinical trial. “I don’t bring it up with patients,” he adds.

Jeffrey Cohen, a neurologist at the Cleveland Clinic in Ohio, is more optimistic. “The evidence already supports using this approach in patients with very active disease that have failed all the available therapies,” he says. But these patients are rare, he adds, and “the available evidence does not support using it more generally.”

Freedman, whose team now considers offering aHSCt to 2 or 3 patients per month, counters that there is a case for broader use. Transplantation should be a treatment option for patients who have rapidly developing and uncontrollable disease⁴, he says. If neurologists use aHSCt only as an absolute last resort, he adds, they might miss the window of opportunity when it offers the most promise. Once the disease becomes too advanced, or patients lose the ability to cope with immuno-ablation, the risk-benefit analysis for aHSCt can become more complex.

More clinical trials could resolve the debate. In 2012, Freedman and his colleagues planned a 114 patient, controlled phase III trial of aHSCt, in which control patients would be randomized to the best available therapy initially and then crossed over onto aHSCt only when their disease progresses⁵. But funding for the trial has remained elusive, and recruitment might also be problematic because people in the control group might drop out to explore other experimental options. Some researchers want to use safer, less toxic drugs to wipe out the immune system in the first phase of the treatment — a move that Freedman says is a mistake because it could leave behind autoimmune cells as well.

Researchers are also advancing an entirely different approach that uses mesenchymal stem cells. Whereas haematopoietic stem cells are intended only to prevent infection in aHSCt, mesenchymal stem cells are thought to offer regenerative capabilities. Preclinical data suggest that mesenchymal cells secrete proteins that can protect axons, improve neuronal survival and induce repair⁶.

A few phase I studies have shown that these cells are safe⁷, and researchers are now looking for signs of efficacy in larger phase II trials. Freedman, for example, is collaborating with colleagues to enrol 40 patients with either relapsing-remitting or secondary progressive MS into a Canadian trial called MESCAMS. The results, which are expected in late 2017 or early 2018, will be pooled with data from another 120 patients who are receiving the same stem-cell treatment in other countries.

Cohen is also set to launch a phase II trial using mesenchymal stem cells in 120 patients with MS in 2017. His group is fine-tuning some technical details, such as how many mesenchymal stem cells to inject into patients, and whether the cells should be frozen before use or injected directly from culture. “These seem like pedestrian issues, but are probably very important,” he says.

Despite the huge amount of uncertainty that surrounds the experimental use of stem cells, a growing number of clinics with dubious scientific qualifications are already cashing in on the stem-cell hype. Profit-driven clinics around the world are offering experimental treatments — that may or may not actually include stem cells — to any patients with MS who will pay. “It’s a big issue,” says Cohen. “It’s hard to know what’s actually going on in these clinics.”

Molson empathizes with desperate people who turn to these clinics through frustration with the slow pace of scientific progress. “I was willing to die to get better. I was in their shoes,” she says. But even she did not properly appreciate the full risks of a legitimate treatment — in a regulated trial, where she knew exactly what she was getting — until it was too late. “People who are thinking about doing this need to talk to their physicians,” she says.

Stem cells may finally be starting to live up to their potential in treating MS — but they are still not ready for prime time. ■

Asher Mullard is a freelance science journalist based in Ottawa, Canada.

1. Atkins, H. L. *et al.* *Lancet* **388**, 576–585 (2016).
2. Sormani, M. P. *et al.* *ECTRIMS Online Library* 146591 (2016).
3. Mullard, A. *Nature Rev. Drug Discov.* **15**, 519–521 (2016).
4. Freedman, M. & Atkins, H. L. *Mult. Scler. J.* **22**, 1258–1259 (2016).
5. Saccardi, R. *et al.* *Mult. Scler.* **18**, 825–834 (2012).
6. Uccelli, A., Laroni, A. & Freedman, M. S. *Lancet Neurol.* **10**, 649–656 (2011).
7. Connick, P. *et al.* *Lancet Neurol.* **11**, 150–156 (2012).

DIET

Changing the recipe

Dietary changes may be able to alleviate the symptoms of multiple sclerosis, but testing the effects of diet will need a different protocol to the one used for drugs.

CATHY CLARKE/MOUNT SINAI HOSPITAL



BY SUJATA GUPTA

Ilana Katz Sand's grocery shopping list for people with multiple sclerosis (MS) includes fish, whole grains, nuts, almond milk, blueberries and carrots. It excludes milk, cheese, meat, crisps and foods containing refined flour or sugar.

Katz Sand, a neurologist and MS specialist at the Mount Sinai Hospital in New York City, says she came up with this diet after scouring the literature for studies, often in animal models, that found a correlation between a given food, supplement or dietary restriction and positive outcomes in MS.

But rather than looking at a single ingredient, or providing patients with ready-made, standardized meals, Katz Sand wants to let patients buy their own ingredients and prepare their food at home. "People are not mice," she says. "We can't put them in a cage and feed them exactly what we want to feed them."

And that's the problem. Hints have begun to emerge that dietary (and lifestyle) changes that reduce inflammation may help people to control their MS, but it is simply not possible to study diet in the same way as drugs. When

researchers test a new drug, they divide people randomly into a treatment or a control group, blind all participants and investigators to that selection, and then provide a standardized amount of the drug or placebo for a set length of time. This highly regimented set-up, known as a double-blind, randomized, controlled trial, is the gold standard for drug research.

But it is almost impossible to standardize diets in this way, and participants always know which group they are in. Those in the control group have little incentive to wait around for months or years, leading to high dropout rates. Efforts are underway to establish a methodology for dietary studies, but it will take time to yield results. Meanwhile, in an environment that effectively favours pharmaceuticals over wellness initiatives, people with MS do not want to wait and are making dietary changes on their own — and it does not always help.

LOW FAT

The idea that people can treat MS by changing their diet has been around for more than half a century. One of the pioneers was neurologist Roy Swank, who, in the 1940s, studied the disease at the University of Montreal in Canada.

He observed that MS was prevalent in the meat-eating countries of Europe and the United States, but was nearly absent in most of Asia, Africa and South America, where meat consumption was much lower. He found the perfect test case in Norway, where people living inland, who ate lots of meat and potatoes, had higher rates of MS than coastal residents, whose diet was based on fish. He sent a questionnaire to neurologists and hospital administrators in both regions and discovered that inland Norwegians were eight times more likely to develop MS than those on the coast.

In 1949, Swank put Norwegians with MS on a diet that involved eating less meat and milk products, but more fish. Other approved foods included a small amount of lean meat, eggs, nuts, cereal and skimmed milk. Swank studied 144 of the participants for several decades. In a 1990 analysis¹, he reported that when patients adhered to a diet of less than 20 grams of saturated fat per day, 95% of them were able to maintain normal activities for the next 30 years. When they deviated from the

Ilana Katz Sand (right) is trying to see whether dietary changes can benefit her patients.

diet, even after years of compliance, the disease invariably got worse.

But Swank's research was flawed by today's standards. When investigators for Cochrane Reviews, systematic reviews of primary research, looked at the purported link between diet and MS in 2012, they determined² that Swank's studies did not meet the inclusion criteria, says lead reviewer Mariangela Farinotti, a lab technician at the Carlo Besta Neurological Institute in Milan, Italy. Swank's studies lacked a control group and stringent screening criteria — and because imaging technologies were not available at the time, it is possible that some of Swank's patients did not have the disease at all. And crucially, during the fourth year of Swank's study, a large number of patients dropped out.

Even worse, Farinotti's team found that none of the studies included in the review showed an alleviation of MS symptoms. "We have no evidence about any kind of dietary treatment being effective on MS," Farinotti says.

VITAL VITAMINS

Despite this lack of evidence, many people with MS have latched onto the idea that lifestyle and dietary changes can help. Studies from Europe, Australia and the United States show that 50–75% of people with MS couple their pharmaceutical treatments with some form of complementary or alternative medicine.

Inflammation in MS occurs when the immune cells attack the nervous system, so any dietary and lifestyle changes known to reduce inflammation do make sense, at least in theory.

And the Cochrane review does not disprove the idea of dietary benefits for people with MS. Farinotti and her team excluded studies of vitamin D, for example, which has been linked to MS, as they were covered under a different review that showed some benefit of vitamin D, but in a study too small to justify making any recommendations. Moreover, because Farinotti's team limited their review to randomized, controlled trials, they examined only six studies, all investigating polyunsaturated fatty acids (PUFAs), which are found mainly in fish. No vitamin, antioxidant or holistic dietary studies made the cut.

This constraint makes some advocates of dietary therapy view the Cochrane review with suspicion. "We don't eat supplements, we eat food," says Terry Wahls, director of the Therapeutic Lifestyle Clinic at the Iowa City VA Hospital. Wahls, who has MS, has created an enormously popular diet for MS that is a variation on the low-grain, meat-heavy Palaeolithic diet, but is higher in vegetables and lower in meat.

Since the Cochrane review was completed, many studies, including some that would have met the reviewers' stringent criteria for inclusion, have shown promise in treating MS with dietary changes. Perhaps most striking is a series of studies that support the Cochrane review in debunking the belief that PUFAs found in fish oil pills can benefit people with MS, instead

attributing those benefits to vitamin D. In 2012, a two-year Norwegian trial³ involving 92 patients, in which half received fish-oil capsules and half were given a placebo, found no difference between the groups in terms of lesion number observed by magnetic resonance imaging (MRI), disease progression, fatigue levels or other measures of quality of life.

Earlier this year, researchers reported results of a study⁴ in which 30 people with relapsing–remitting MS, a form that affects 85% of those initially diagnosed with the disease, received high doses of vitamin D. They were slightly less likely to experience relapses, and had fewer lesions visible on MRI scans, than the 23 control subjects.

Both vitamin D and PUFAs are high in certain species of fish, says Øivind Torkildsen, a neurologist at the University of Bergen in Norway, who was involved with both studies. "We were looking at the wrong compound."

Another study published this year⁵ — a randomized, controlled trial of 61 patients — found that those who ate no animal products or took no supplements for a year (the McDougall diet, named after John McDougall, a doctor inspired by Swank) had lower levels of fatigue than those in the control group.

SMART DESIGN

These studies are starting to show that dietary changes can benefit patients with MS, but designing trials to test diets remains enormously challenging. Wahls had to make some difficult choices as she designed a trial to compare her high-fat, modified Palaeolithic diet with Swank's low-fat diet. The study, which is funded by a US\$1 million grant from the US National Multiple Sclerosis Society, could shed light on the rift in the MS community over whether fat is good or bad for people with MS.

Wahls will observe 100 people with relapsing–remitting MS for 12 weeks before splitting them randomly into the Swank or the Wahls treatment. For 12 weeks, participants will receive dietary guidance for their allotted diet, followed by 12 weeks for them to follow the diet on their own. Wahls and her team will then use a questionnaire to measure fatigue levels. Because fatigue changes faster than mental clarity and gait, Wahls says, it is easier to study in the short term than other measures. To minimize drop-out, she opted against using a control group and will instead compare fatigue levels between the observation and treatment periods.

But George Jelinek, a neuroepidemiologist at the University of Melbourne in Australia, who like Wahls began studying the link between diet and MS after being diagnosed with the disease, thinks that the lack of a control group is a problem. "What if the diets are equivalent?" he asks. "It could be a placebo effect for both of them."

Even if the treatments both work equally

well, the 12 week observation period serves as a proxy for a control group, Wahls says. "The best outcome," she says, "is that both diets are relatively equivalent, and they both do a greater job of reducing fatigue than during the observation period."

Meanwhile, at Mount Sinai, Katz Sand was grappling with the same issues as her colleagues as she sought to set up her own study. How was she going to measure participants' adherence to the diet, for instance? Should she jump on the high-fat or low-fat bandwagon? How could she stop people dropping out?

To sort out her methodology, Katz Sand designed a 30 person, 6 month pilot study that launched at the start of October. After some preliminary research, she opted to test a modified Mediterranean diet that is rich in whole grains, fruits, vegetables and healthy fats, such as those found in fish, nuts and avocados. She excluded foods that contain saturated fat, such as meat and cheese. She allowed vitamin D supplements, but only if participants fell below the established normal range.

People in the treatment group also receive intensive education about the diet and have regular meetings with a nutritionist. To measure adherence, participants must complete dietary questionnaires and have routine blood tests to measure their levels of salt, fatty acids and other nutrients. To discourage participants from dropping out, those in the control group will meet every month to discuss health and wellness issues in MS. Katz Sand has also implemented a modified waiting list control group, so that people assigned to the control group will be offered the same treatment (apart from the regular meetings and monitoring) as the test group at the end of the study.

The difficulties of studying wellness initiatives, such as changes to diet and lifestyle, affect the assessment of treatments for a wide range of conditions. But MS provides a measurable way to study the effect of diet on long-term health. Nicholas LaRocca, vice-president of health-care delivery and policy research at the National Multiple Sclerosis Society, hopes investigators will one day be able to show quantitatively how dietary changes alleviate the symptoms of MS.

In the meantime, if Katz Sand's methodology works, she will have helped to contribute to the development of a new gold standard for clinical trials that will let wellness-based treatments stand on an equal footing with their pharmaceutical counterparts. ■

Sujata Gupta is a freelance science writer based in Burlington, Vermont, USA.

1. Swank, R. L. & Dugan, B. B. *Lancet* **336**, 37–39 (1990).
2. Farinotti, M. et al. *Cochrane Database Syst. Rev.* **12**, CD004192 (2012).
3. Torkildsen, Ø. et al. *Arch. Neurol.* **69**, 1044–1051 (2012).
4. Muris, A.-H. *ECTRIMS Online Library* 146467 (2016).
5. Yadav, V. et al. *Mult. Scler. Relat. Disord.* **9**, 80–90 (2016).

Myeloid cells — targets of medication in multiple sclerosis

Manoj K. Mishra and V. Wee Yong

Abstract | Discussions of multiple sclerosis (MS) pathophysiology tend to focus on T cells and B cells of the adaptive immune response. The innate immune system is less commonly considered in this context, although dendritic cells, monocytes, macrophages and microglia — collectively referred to as myeloid cells — have prominent roles in MS pathogenesis. These populations of myeloid cells function as antigen-presenting cells and effector cells in neuroinflammation. Furthermore, a vicious cycle of interactions between T cells and myeloid cells exacerbates pathology. Several disease-modifying therapies are now available to treat MS, and insights into their mechanisms of action have largely focused on the adaptive immune system, but these therapies also have important effects on myeloid cells. In this Review, we discuss the evidence for the roles of myeloid cells in MS and the experimental autoimmune encephalomyelitis model of MS, and consider how interactions between myeloid cells and T cells and/or B cells promote MS pathology. Finally, we discuss the direct and indirect effects of existing MS medications on myeloid cells.

Multiple sclerosis (MS) is an immune-mediated disease of the CNS. The involvement of lymphocytes of the adaptive immune system in MS pathogenesis is widely acknowledged, whereas innate myeloid antigen-presenting cells (APCs), which include monocytes, macrophages, dendritic cells and microglia, are less commonly linked to the disease. This discrepancy is surprising, as macrophages and microglia are the predominant inflammatory cells in active or chronic MS plaques^{1–3}, and they persist in secondary progressive MS². Activated myeloid APCs — referred to as myeloid cells in this Review to emphasize their functions beyond antigen presentation — can produce several inflammatory cytokines, free radicals, proteases and other mediators of tissue injury in MS. Moreover, considerable interactions between lymphocytes and myeloid cells augment the inflammatory milieu, the components of which in turn destroy oligodendrocytes and cause axonal injury and demyelination⁴.

In this Review, we introduce the types, subtypes and functions of myeloid cells, discuss their involvement in MS and the inflammatory animal model experimental autoimmune encephalomyelitis (EAE), consider the mechanisms by which they cause tissue injury, and discuss the impact on myeloid cells of immunomodulators that are used to treat MS.

Myeloid cells

Myeloid cells include monocytes, macrophages, microglia and dendritic cells. Each of these cell types has distinct developmental origins, and serves different

functions at different sites. Furthermore, subtypes of each group are defined according to their functions and expression of cell surface receptors.

Origins

The origins of myeloid cells have been extensively discussed previously^{5–7}, so will only be summarized briefly here (FIG. 1). During gestational development, microglial and macrophage precursors develop from erythromyeloid progenitors in the yolk sac and migrate into their target tissues as soon as the circulatory system is established^{6,8–10}. These precursors give rise to CNS-resident microglia and tissue-resident macrophages, such as Kupffer cells in the liver and Langerhan cells in the skin. The overall development of these cells depends on their expression of the PU.1 transcription factor and macrophage colony-stimulating factor 1 receptor (CSF-1R).

After birth, microglia are not thought to be replenished by extra-CNS cells^{6,11}, whereas monocytes continue to be produced and maintained by haematopoietic stem cells (HSCs) within the bone marrow. Conventional dendritic cells originate from myeloid-derived precursors in the bone marrow, but dendritic cells can also be derived from monocytes under inflammatory conditions.

Functions and subtypes

Monocytes. Circulating monocytes are classified into two groups. Monocytes that express CD115, lymphocyte antigen 6C (Ly6C) and C-C chemokine receptor

Hotchkiss Brain Institute and the Department of Clinical Neurosciences, University of Calgary, 3330 Hospital Drive, Calgary, Alberta, T2N 4N1, Canada.

Correspondence to V.W.Y. vyong@ucalgary.ca

doi:10.1038/nrneuro.2016.110
Published online 12 Aug 2016

Key points

- The roles of lymphocytes in multiple sclerosis (MS) pathophysiology are well known, but myeloid cells, including monocytes, macrophages, microglia and dendritic cells, are also important
- Interactions between lymphocytes and myeloid cells exacerbate injurious processes
- Myeloid cells are not the main targets of immunomodulators that are used to treat MS, but are still affected by them in disease-relevant ways
- The actions of MS immunomodulators on myeloid cells contribute to the clinical efficacy of these therapeutic approaches

type 2 (CCR2) on their surface are known as immature proinflammatory monocytes^{5,12}. These cells home in on an area of injury, where the local tissue milieu promotes their differentiation into proinflammatory macrophages¹³. The second subset of circulating monocytes express low levels of Ly6C and high levels of CX3CR1, and do not express CCR2. These cells are known as regulatory, patrolling monocytes, and can become tissue-resident macrophages under healthy conditions¹². These patrolling cells can include perivascular macrophages that are found in the perivascular space between the endothelial and parenchymal basement membranes of post-capillary venules, where leukocytes gather before migrating across the blood–brain barrier^{14,15}. Perivascular macrophages can detect antigens in the perivascular space and help to activate lymphocytes, which subsequently enter the CNS parenchyma. Evidence for this mechanism comes from the observation that depletion of monocytes and macrophages by clodronate-containing liposomes blocks the invasion of the CNS parenchyma by leukocytes in animals with EAE¹⁶. Other macrophage populations in the non-diseased brain include meningeal macrophages and choroid plexus macrophages; the latter detect disturbances in the brain–cerebrospinal fluid barrier that result from injury or infection.

Macrophages and microglia. Like monocytes, macrophages and microglia can be subclassified on the basis of their polarization¹⁷. For many years, two major subsets of macrophages have been defined: ‘M1’ macrophages (classical activation) and ‘M2’ macrophages (alternative activation)^{12,17}. IFN- γ and lipopolysaccharide are activators of M1 macrophages. On activation, these cells generate high levels of nitric oxide by producing inducible nitric oxide synthase, which uses arginine as a substrate. M2 macrophages are activated by local tissue damage or parasitic infestation. In the commonly studied ‘M2a’ phenotype, which is generated by IL-4 and IL-13 treatment in tissue culture studies, arginine is used as a substrate for arginase-1 to form ornithine, a precursor of polyamines, proline and collagen, which are needed for wound healing¹⁷. M1 and M2 macrophages also produce different cytokines (FIG. 2): M1 macrophages produce high levels of IL-1 β , IL-12 and tumour necrosis factor (TNF), which are largely proinflammatory, whereas M2 macrophages secrete predominantly regulatory and/or anti-inflammatory cytokines, including IL-4 receptor antagonist, IL-10 receptor antagonist and IL-1 receptor antagonist.

The M1–M2 subdivision of macrophages emphasizes the functional differences, but the classification is now recognized as being oversimplified. Macrophages with a variety of different markers and properties can develop, depending on the activation stimuli and micro-environmental influences¹⁸. A 2014 study in which human macrophages were activated with different stimuli revealed a spectrum of at least nine different varieties on the basis of transcriptome profiles¹⁹. Better delineation of the subtypes of macrophages — and probably of microglia and other myeloid cells — is needed²⁰. For the remainder of this Review, we will refer to M1 and M2 macrophages as proinflammatory and regulatory, respectively, except where cells were comprehensively characterized in specific studies. However, the current consensus is that proinflammatory and regulatory macrophages are just two varieties in a spectrum of many macrophage phenotypes (FIG. 2).

Microglia are distinguishable from bone marrow-derived macrophages, as they express lower levels of CD45²¹. Under basal conditions, microglia are highly dynamic and constantly patrol the CNS microenvironment²². Activation of these cells leads to an increase in their size, upregulated expression of several molecules, and initiation of effector functions, including antigen presentation, chemoattraction, and phagocytosis of debris or toxic factors such as ferrous ions. Activated microglia can be injurious to the CNS, or provide benefits such as neuroprotection, mobilization of neural precursors, remyelination and axonal regeneration, a dichotomy that has previously been discussed extensively^{23–26}.

Dendritic cells. Conventional dendritic cells are important myeloid cells that activate T cells both in the healthy state and in response to pathogenic triggers. These cells express CD11b and CD11c, and are distinct from B220⁺CD11c⁺ plasmacytoid dendritic cells²⁷, which have mixed characteristics of lymphocytes and dendritic cells. Plasmacytoid dendritic cells are thought to be less efficient than conventional dendritic cells at presenting antigens to activate T cells²⁸, but evidence shows that they have complex roles. In one study, adoptive transfer of myelin antigen-loaded plasmacytoid dendritic cells into mice with ongoing EAE led to recovery from the condition²⁹.

Myeloid cells in EAE

Two of the most consistent aberrations observed in EAE and MS are the recruitment of blood-borne monocytes to become CNS macrophages, and the activation of resident microglia (FIG. 3). In this section, we discuss studies that have identified these aberrations and the underlying mechanisms.

Monocytes and macrophages

In EAE, egress of proinflammatory monocytes from the bone marrow increases soon after immunization of mice with myelin peptides to induce disease, and these monocytes accumulate in the blood before entering the CNS^{30,31}. Substantial transmigration of monocytes into the spinal cord coincides with the clinical onset of

disease^{30,32} and, as demonstrated by the use of irradiation and parabiosis in combination, infiltration of monocytes into the CNS is necessary for the clinical severity of EAE to worsen³³. The infiltrating monocytes were found to trigger progression of EAE disability, but did not contribute to the microglial pool in the CNS.

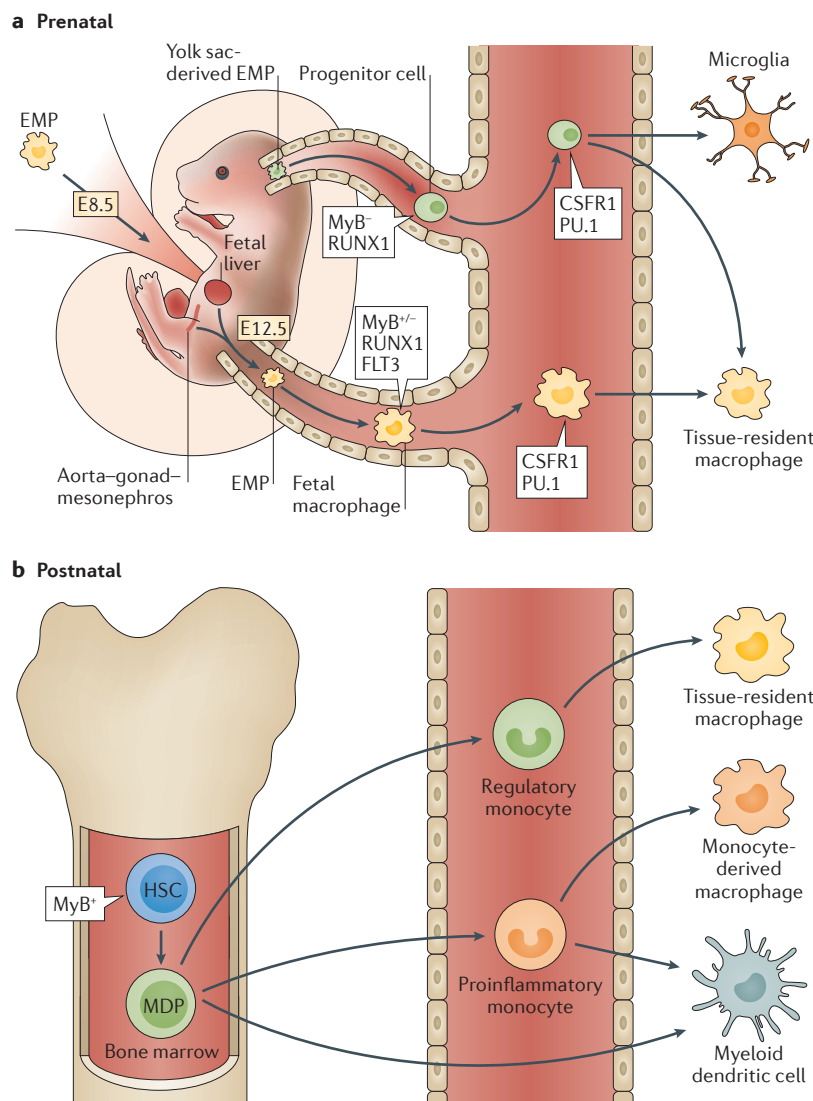


Figure 1 | Development of myeloid cells. a | During gestational development of the mouse, after embryonic day 8.5 (E8.5), yolk sac-derived erythromyeloid progenitor cells (EMPs) migrate through a nascent circulatory system to the brain and other tissues to become long-lived microglia and tissue-resident specialized macrophages (top). After embryonic day 12.5 (E12.5), other fetal macrophage populations derived from EMPs, such as those in the aorta-gonad-mesonephros and liver, mature into tissue-resident macrophages. Maturation of different macrophage populations depends on expression of the different transcription factors shown. **b** | Postnatally and throughout life, monocytes are produced by haematopoietic stem cells (HSCs) and myeloid-derived precursors (MDPs), which also give rise to dendritic cells. At least two populations of circulating monocytes exist: regulatory monocytes that are available to become resident, tissue-specific macrophages, and proinflammatory monocytes that migrate into tissues where the local milieu promotes their differentiation into proinflammatory macrophages. CSFR1, colony-stimulating factor 1 receptor; FLT3, Fms-like tyrosine kinase 3; MyB, transcriptional activator Myb; PU.1, transcription factor PU.1; RUNX1, runt-related transcription factor 1.

Transmigration of leukocytes, including monocytes, across the glia limitans and into the CNS parenchyma³⁴ requires matrix metalloproteinases (MMPs); monocytes, macrophages and microglia express a variety of MMPs in EAE and when activated *in vitro*^{35,36}. Furthermore, levels of extracellular MMP inducer (CD147), an upstream regulator of MMP expression, are elevated in leukocytes (including monocytes) and on lesional macrophages and microglia in EAE. Inhibition of CD147 activity reduces EAE severity²¹, indicating that expression of CD147 and MMPs by myeloid cells contributes to disease progression.

Multiple lines of evidence indicate that proinflammatory monocytes are important in EAE. For example, in mice, numbers of proinflammatory monocytes increase in the blood within 1 day of myelin peptide immunization to induce EAE³². Depletion of monocytes and/or macrophages in rodents, either by administration of agents such as silica dust or clodronate liposomes or by CCR knockout³¹, reduces or prevents the clinical manifestations of EAE^{37,38}. Similarly, depletion of monocytes and/or macrophages attenuates influx of CD4⁺ T cells into the CNS^{16,39}. Furthermore, medications, including dipyridamole⁴⁰, that inhibit the activation of proinflammatory myeloid cells ameliorate EAE in mice, and administration of regulatory monocytes has the same outcome^{41,42}.

Besides antigen presentation that activates T cells, proinflammatory myeloid cells have numerous effector functions in EAE, including the production of cytokines and chemokines that can further promote inflammatory responses. These responses include recruitment of lymphocytes into the CNS^{30,43}, phagocytosis of debris (degenerated myelin, for example)⁴⁴, and generation of oxidative stress and other mediators of injury^{43,45–47}. For example, through the use of serial block-face electron microscopy of fluorescently labelled microglia, monocytes and macrophages from mice with EAE, monocyte-derived macrophages were shown to associate with the nodes of Ranvier and initiate demyelination, and the resultant debris was microcytosed by microglia⁴⁸.

Evidence shows that another effector function of macrophages that is relevant to EAE is the linking of sympathetic stress with the inflammatory axis. Noradrenaline — an important mediator of the response to many stressors — is known to be involved in the development of EAE⁴⁹, and one study has shown that macrophages contribute to the underlying mechanism: increased production of noradrenaline by macrophages owing to inactivation of the transcription factor Nr4a1 promoted recruitment of leukocytes into the CNS and exacerbated EAE⁵⁰. The same study also showed that myeloid-specific deletion of tyrosine hydroxylase, the rate-limiting enzyme in noradrenaline synthesis, protected against EAE in mice. Interestingly, compared with healthy controls, patients with MS were found to express higher levels of tyrosine hydroxylase in their CD14⁺ monocytes⁵⁰.

Microglia

Microglial activation is a hallmark of demyelinating lesions in EAE^{51,52}. Moreover, in the chronic phase of EAE, when the density of T cells in lesions has subsided,

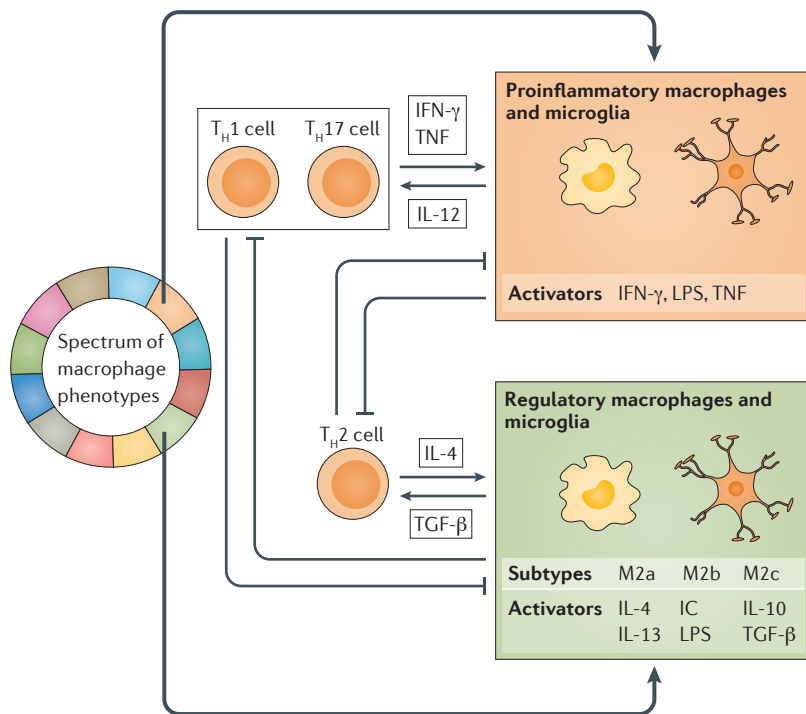


Figure 2 | Characteristics of macrophage subsets. Subsets of macrophages, and probably microglia, include proinflammatory and regulatory cell types that are generated by specific activators and produce specific cytokines. Facilitatory interactions exist between T_H1 or T_H17 cells and proinflammatory macrophages, and between T_H2 cells and regulatory macrophages. This classification is likely to undergo further revision, as the proinflammatory and regulatory subsets are now thought to represent only two examples from a spectrum of many potential phenotypes (left). IC, immune complex; LPS, lipopolysaccharide; TGF- β , transforming growth factor β ; T_H , T helper; TNF, tumour necrosis factor.

accumulation of activated microglia persists in areas of neuronal dysfunction⁵². Activated microglia can damage oligodendrocytes and neurons by producing free radicals and proinflammatory cytokines, such as TNF⁵³. Microglia can also boost activation of T cells that have entered the CNS⁵⁴. In tissue culture, microglia that had been preactivated with IFN- γ were nearly as efficient as dendritic cells at inducing proliferation of naive T cells and restimulating proinflammatory T helper 1 (T_H1) cells, whereas astrocytes were ineffective; conversely, all three cell types were capable of restimulating regulatory T_H2 cells⁵⁵. Moreover, activated T cells can engage microglia to stimulate production of several cytokines and other molecules that can further promote injury⁴.

Evidence shows that in early lesion formation in EAE, microglia form perivascular clusters before demyelination or clinical signs become apparent⁵⁶. Microglial activation and production of reactive free radicals are thought to be initiated by the serum protein fibrinogen, which leaks across the disrupted blood–brain barrier. Depletion of microglia⁵⁷ or downregulation of genes within microglia⁵⁸ results in attenuation of disease severity and demyelination in EAE.

In other models of neurological injury, microglia have been shown to have protective roles that might be relevant in EAE. In a model of stroke, for example,

microglia reduced toxicity by engulfing neutrophils⁵⁹. 3D electron microscopy has shown that microglia that have been activated by a systemic lipopolysaccharide injection disrupt inhibitory GABAergic presynaptic terminals in projections from cortical neurons in adult mice, resulting in elevated synchronized firing of cortical neurons, expression of neurotrophic molecules, and decreased neuronal apoptosis⁶⁰. Although persistent activation of microglia results in detrimental outcomes, the protective effects in models of other disease suggest that these cells could serve useful functions at certain stages of EAE disease, and this possibility remains to be clearly defined.

Myeloid cells in multiple sclerosis

Activated macrophages and microglia are hallmarks of active lesions with ongoing demyelination and axonal injury in MS^{1–3,61}, and several lines of evidence indicate that these cells are associated with brain injury. ¹¹C-PK11195 PET imaging of microglia shows that microglial activity in the cortical grey matter of living individuals with MS corresponds with progression of disability⁶². Indeed, microglial activation can be detected before demyelination in ‘preactive’ lesions in MS. Histological examination of autopsy specimens has revealed that in normal-appearing white matter from individuals with MS, the extent of local microglial inflammation correlates with the extent of paranodal disruption³. Another study showed that cortical demyelination and diffuse axonal injury in normal-appearing white matter was associated with profound microglial activation⁶³. Furthermore, microglial activation can be detected in normal-appearing white matter before leukocyte infiltration, demyelination or blood–brain barrier disruption^{64,65}. Among patients with MS, those who exhibit prominent microglial activation are the same patients that exhibit signs of cortical lesions⁶⁶. In progressive MS, nodules that represent aggregates of microglial cells are very prominent in lesions². These findings all suggest that inhibition of neurotoxic microglia is likely to be a useful therapeutic strategy in MS, but many currently used MS medications have a poor capacity to cross the blood–brain barrier.

Evidence from one study suggests that markers associated with proinflammatory (including CD40⁺, CD86⁺ and CD32⁺) and regulatory (including mannose receptor-expressing and CD163⁺) cells are expressed throughout the course of an active demyelinating MS lesion by activated macrophages and microglia that have engulfed myelin⁶⁵. This finding led to the conclusion that although activated macrophages in active MS lesions predominantly exhibit proinflammatory characteristics, a considerable number of these cells have a mixed proinflammatory and regulatory activation status. Similarly, in a study of early MS lesions characterized by microglial nodules in the normal-appearing white matter, the microglial populations exhibited both proinflammatory and anti-inflammatory phenotypes⁶⁷. In a third study, activated microglia in preactive and remyelinating lesions in MS brain specimens were also found to express mixed ‘M1’ and ‘M2’ markers⁶⁸; this study also showed that, at least in tissue culture,

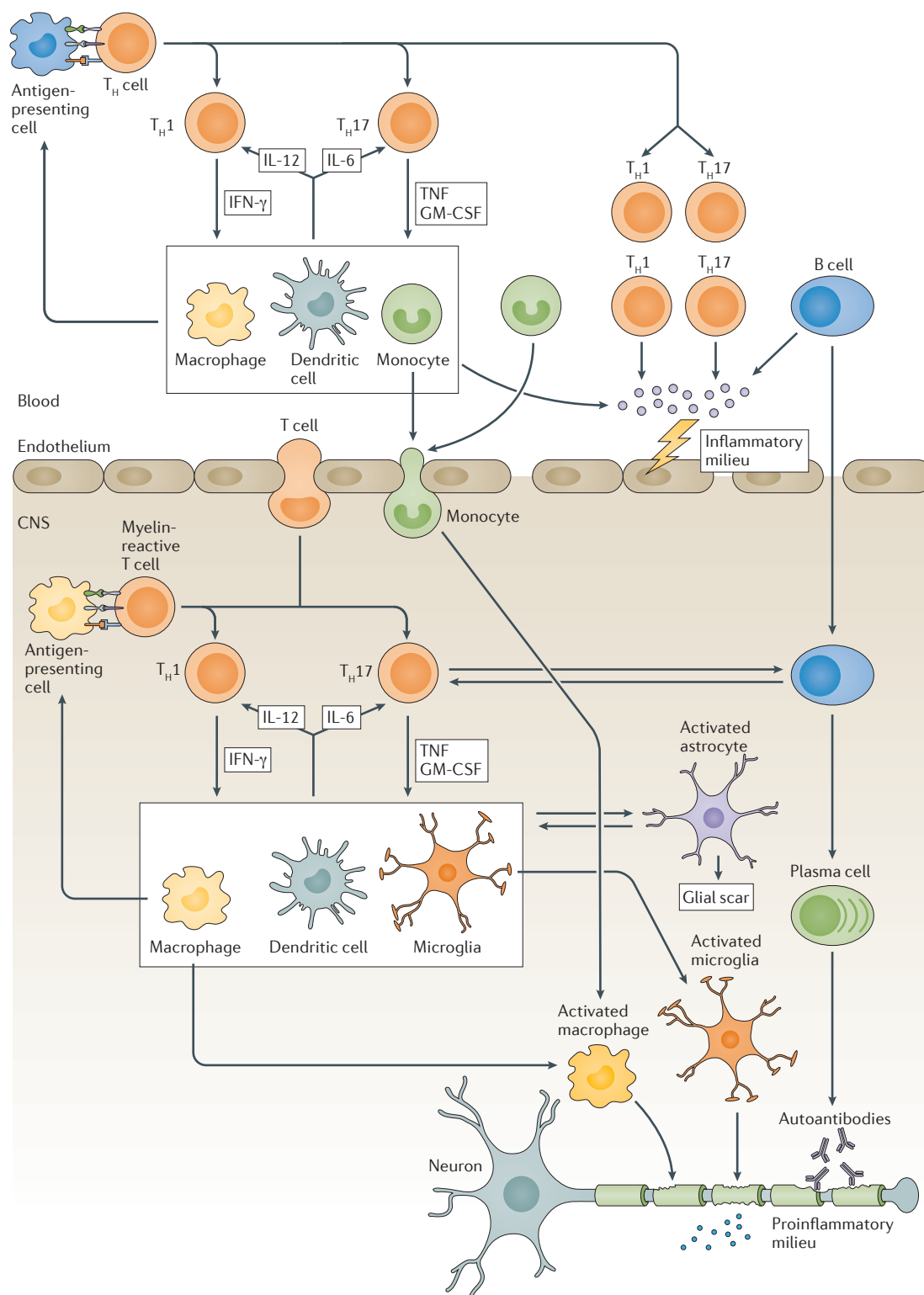


Figure 3 | Postulated immune cascades in multiple sclerosis. Myeloid cells and B cells are competent antigen-presenting cells that are thought to present a molecular mimic of a myelin peptide to generate myelin-reactive T_H1 and T_H17 cells that can further activate myeloid cells. Activated immune cells produce an inflammatory milieu (proinflammatory cytokines, chemokines, proteases, free radicals and glutamate) that damages the endothelium; the same cells can then adhere to endothelial cells and produce proteases, such as the matrix metalloproteinases, that enable them to transmigrate across the blood–brain barrier into the CNS. Local antigen presentation in the CNS can further activate T cells, leading to cycles of activation of myeloid cells and B cells in the CNS. Activated immune cells and the proinflammatory milieu are thought to contribute to neural injury. GM-CSF, granulocyte macrophage colony-stimulating factor; T_H, T helper; TNF, tumour necrosis factor.

microglia that were polarized in one direction could convert to the opposite polarization when exposed to the counter polarization cytokine regimen.

Similar associations with MS disease activity have been reported for myeloid cells in the circulation. Enzyme-linked immunospot assays have detected higher levels of monocytes that secrete proinflammatory IL-6 and IL-12 in untreated patients with MS than in healthy controls⁶⁹. Assays of dendritic cells have shown similar differences in levels of cells that secrete IFN- γ , TNF and IL-6 (REF. 70). Elevation in the numbers of proinflammatory monocytes that produce IL-12 correlates with the presence of gadolinium-enhancing MRI lesions in patients with MS⁷¹. A recent study has also indicated that the numbers of CD16⁺ monocytes, which are presumed to have regulatory roles, are relevant to MS⁷². These cells were found to be depleted in the circulation of patients with MS; within the cerebrospinal fluid, the ratio of regulatory CD16⁺ to proinflammatory CD16⁻ monocytes was lower in patients with relapsing–remitting MS than in controls with no inflammation.

Myeloid cells in mice and humans have some differential properties: expression of markers for proinflammatory or regulatory subsets differs (for example, Ly6C is a proinflammatory marker of mouse but not human cells, which are better identified according to their expression of CD14 and CD16), and a mixed phenotype seems to be more common in humans than in mice. Nevertheless, evidence suggests that the properties of myeloid cells in mice generally predict the properties of their human counterparts.

Myeloid and T cell interaction

Dendritic cells, macrophages and microglia can present antigens and, consequently, activate and polarize T cells (FIG. 3). CNS sites at which myeloid cells and T cells interact include the CNS parenchyma after leukocyte infiltration; the perivascular space between the endothelial and parenchymal basement membranes of post-capillary venules, where leukocytes gather before migrating across the blood–brain barrier^{14,15}; and the subarachnoid space of the leptomeninges⁷³. Molecules from differentiated T_H1 and T_H17 cells can, in turn, induce further myeloid cells to become proinflammatory. This vicious cycle damages the blood–brain barrier and facilitates the transmigration of activated T cells and other leukocytes into the CNS. Within the CNS, T cells can interact with microglia or other infiltrating myeloid cells, resulting in mutual reactivation^{74,75}, and propagation of neural injury (FIG. 3).

The interaction between T cells and myeloid cells seems to be mediated by the cytokine GM-CSF (granulocyte macrophage colony-stimulating factor), which is a requisite for the development of EAE^{76–78}. GM-CSF produced by encephalitogenic T cells in EAE causes expression of chemokines by CCR2⁺Ly6C^{high} monocytes and activation of inflammasomes in the same cells, leading to sustained inflammation and CNS tissue injury⁷⁴. Interestingly, deletion of the GM-CSF receptor gene (*Csf2rb*) in conventional or monocyte-derived dendritic cells or microglia did not improve the

Figure 4 | Proposed actions of disease-modifying multiple sclerosis therapies on myeloid cells. In the periphery, glatiramer acetate and laquinimod promote the generation of regulatory myeloid cells. Dimethylfumarate and IFN- β reduce generation of proinflammatory T_H1 and T_H17 cells by affecting myeloid cell antigen presentation. The availability of proinflammatory cells can be reduced through depletion of B cells with rituximab or ocrelizumab, depletion of T_H cells, B cells and monocytes with alemtuzumab, or inhibition of lymphocyte proliferation with teriflunomide. Dimethylfumarate, fingolimod and laquinimod can cross the blood–brain barrier and are thought to inhibit microglia activity in the CNS. Regulatory myeloid cells and T_H2 cells generated by glatiramer acetate in the periphery also modulate microglial activity in the CNS. T_H1 and T_H17 cell production as a result of antigen presentation within the CNS might also be affected by laquinimod and glatiramer acetate-generated regulatory myeloid cells. These effects in the CNS probably reduce neurotoxic effects in multiple sclerosis. MMP, matrix metalloproteinase; T_H, T helper; VLA-4, very late antigen-4; VCAM-1, vascular cell adhesion protein 1.

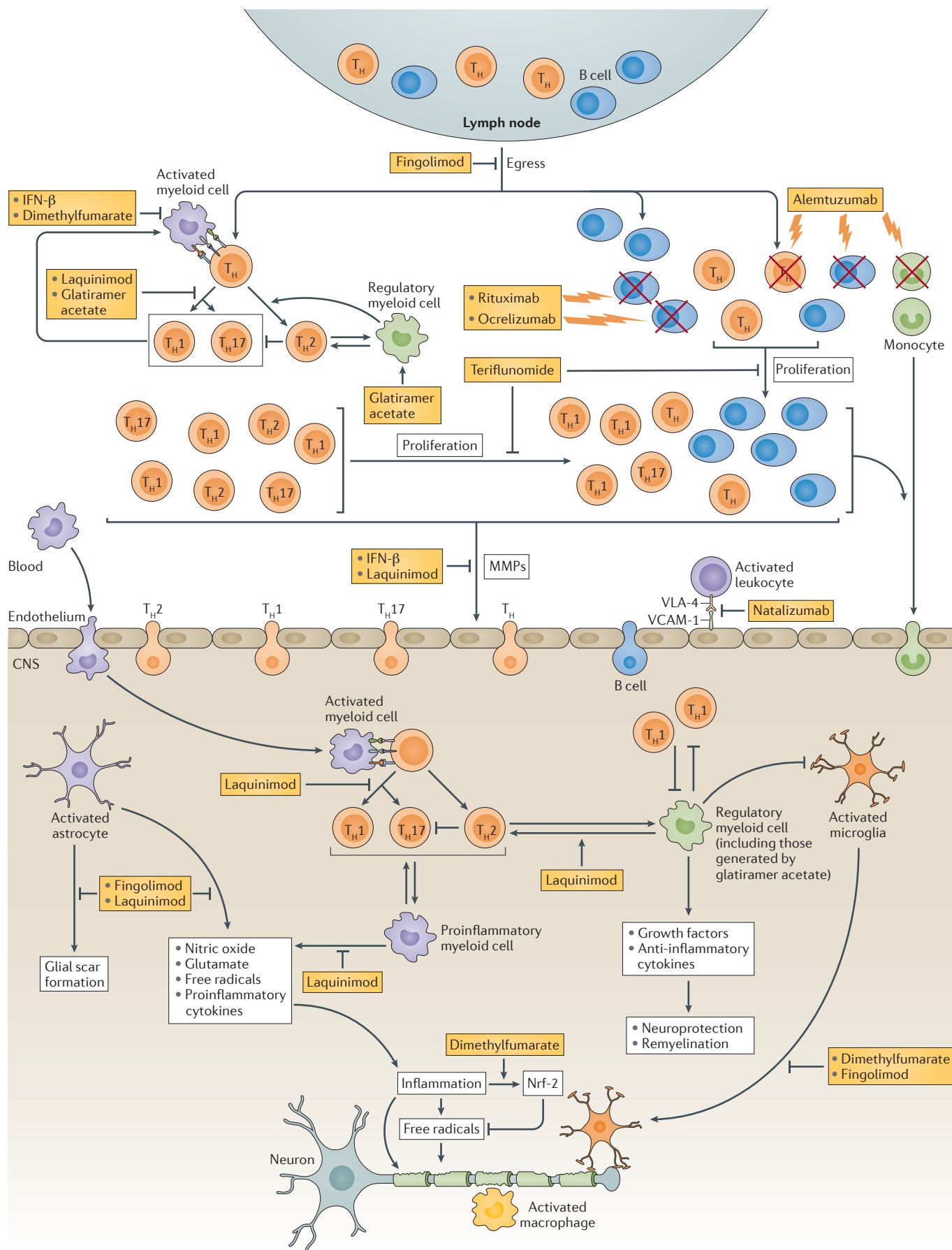
outcomes of EAE, whereas loss of GM-CSF signalling in CCR2⁺Ly6C^{high} monocytes led to a phenocopy of the resistance to EAE found in *Csf2rb*-deficient mice⁷⁴. In a proposed model of the pathogenic mechanisms, GM-CSF produced by T_H cells coordinates the recruitment of CCR2⁺Ly6C^{high} monocytes into the CNS and, consequently, promotes tissue injury⁷⁹.

Action of MS drugs on myeloid cells

T cells and B cells are most frequently discussed as targets of disease-modifying therapies in MS^{80,81}, but these drugs also affect myeloid cells. In this section, we describe the direct and indirect effects of MS medications on myeloid cells (FIG. 4). A recent review comprehensively summarized the biology of myeloid cells as potential drug targets, including the development of antagonists to receptors on myeloid cells, such as CCR2, P2X7 and CSF1 (REF. 82). Our focus, however, is on whether emerging or established MS medications act on myeloid cells even if these cells were not the original intended targets.

IFN- β

IFN- β has diverse mechanisms of action in MS⁸³, and several actions on myeloid cells. For example, in cells obtained from patients with MS or healthy controls, IFN- β alters the expression of co-stimulatory molecules that are needed for antigen presentation on myeloid cells^{84,85}, and inhibits dendritic-cell-mediated T_H17 cell differentiation⁸⁶. In mature dendritic cells, IFN- β inhibits CCR7 expression and MMP-9 production, and reduces the capacity of these cells to migrate into the CNS, thereby potentially preventing the reactivation of encephalitogenic T cells within the CNS⁸⁷. Similarly, exposure of cultured monocytes from patients with MS to IFN- β reduces their production of various MMPs but increases expression of physiological MMP inhibitors known as TIMPs; these changes indicate a less migratory phenotype⁸⁸. Finally, in monocytes and dendritic



cells from patients with MS, IFN- β upregulates the expression of B7-H1, a strong inhibitor of CD4⁺ T-cell activation⁸⁹.

Genome-wide expression profiling and functional studies in patients with MS have determined that IFN- β non-responders have higher levels of co-stimulatory molecules, such as CD86, on myeloid dendritic cells before the initiation of treatment with IFN- β ⁹⁰. The finding suggests that patients with MS who have more preactivated dendritic cells are less likely to respond to IFN- β therapy. IFN- β treatment also promotes apoptotic responses in mature dendritic cells⁹¹, providing further evidence that the drug effects myeloid cells.

Studies in EAE have also demonstrated the impact of IFN- β on myeloid cells. Studies of mice with defective IFN- β signalling have indicated that the therapeutic effect of IFN- β is attributable to its ability to stimulate production of the regulatory cytokine IL-27 — which limits T_H17 cell development — by myeloid cells⁹². Furthermore, mice that lack the widely expressed type 1 interferon receptor develop more-severe EAE than do wild-type mice, and this increase in severity was replicated in mice with selective ablation of type 1 interferon receptor in myeloid cells; by contrast, deletion of the type 1 interferon receptor in B cells or T cells did not lead to an increase in disease severity⁹³. These results strongly suggest that targeting of myeloid cells by IFN- β is involved in the amelioration of neuroinflammation.

Experiments in cells from patients with MS provide further evidence that myeloid cells are targets of IFN- β . Specifically, monocytes from IFN- β -treated patients produce lower levels of proinflammatory cytokines, including IL-12p70 (REF. 94), and higher levels of the regulatory cytokine IL-10 (REF. 85) than do monocytes from patients not treated with IFN- β , implying a switch to regulatory myeloid cell subtypes. Monocytes from patients treated with IFN- β also have a decreased capacity to generate reactive oxygen species⁹⁵, but produce high levels of brain-derived neurotrophic factor (BDNF)⁹⁶. Monocytes from individuals treated with IFN- β also express higher than normal levels of immunoglobulin-like transcript 3, a receptor involved in inhibition of immune responses⁹⁷. Furthermore, dendritic cells from patients who receive IFN- β have normal levels of IFN- γ and TNF, whereas levels of these factors are elevated in cells from untreated patients with MS⁷⁰.

Glatiramer acetate

The best-documented mechanism of action of glatiramer acetate is the generation of regulatory, anti-inflammatory T_H2 cells that secrete IL-4, IL-10 and transforming growth factor β , but this drug also affects myeloid cells. Glatiramer acetate leads to the generation of regulatory myeloid cells (referred to as type II APCs in some studies)^{42,98,99}, which, according to several lines of evidence, are responsible for the generation of T_H2 cells⁴². Furthermore, adoptive transfer of regulatory monocytes that have been produced in response to glatiramer acetate can confer protection against EAE on recipient mice⁴².

Evidence indicates that glatiramer acetate treatment of patients with MS reduces expression of the co-stimulatory molecule CD40 on their blood-derived dendritic cells, and this reduction is associated with depletion of T_H1 cells but increased numbers of CD4⁺ T cells expressing regulatory markers¹⁰⁰. Glatiramer acetate has also been shown to inhibit nitric oxide production and to increase the phagocytic activity of microglia and monocytes, further suggesting that the drug can induce polarization of these cells away from a proinflammatory phenotype¹⁰¹. Similarly, exposure of microglia to glatiramer acetate increases their production of IL-10 and their phagocytic activity, but decreases their production of TNF — all features of a regulatory subtype¹⁰².

Evidence that glatiramer acetate acts on myeloid cells has also come from studies of monocytes from the blood of patients with MS who are receiving this therapy. These cells, when compared with those from untreated MS patients or healthy controls, exhibit increased phagocytic activity¹⁰³ and production of the anti-inflammatory soluble IL-1 receptor antagonist¹⁰⁴, and have a regulatory phenotype characterized by decreased production of IL-12 and increased IL-10 (REFS 98,99). Moreover, monocytes isolated from patients with MS 3 days after the initiation of daily glatiramer acetate therapy exhibit increased production of IL-10 on *ex vivo* activation¹⁰⁵.

Together, this evidence shows that glatiramer acetate has many effects on myeloid cells that directly or indirectly promote the development of regulatory phenotypes in these cells and T cells. ¹¹C- PK11195 PET imaging further supports this hypothesis: use of this technique to detect microglia revealed that microglial activation in cortical grey and white matter had decreased at 1 year after exposure to glatiramer acetate compared with pretreatment values¹⁰⁶.

Fingolimod

Fingolimod was the first oral therapy to be approved for the treatment of relapsing–remitting MS. The phosphorylated form of fingolimod is an agonist of sphingosine-1-phosphate (S1P) receptors, and modulates expression of these receptors at the cell surface¹⁰⁷. The S1P receptors are present on various immune cell types and myeloid cells. By interfering with S1P receptor signalling in immune cells, fingolimod sequesters lymphocytes within the lymph node and prevents their access to target sites where they would inflict injury¹⁰⁸.

Myeloid cells are affected by fingolimod, and several lines of evidence indicate that they tend to become less proinflammatory as a result. Expression of S1P receptor 3 is higher in regulatory monocytes than in proinflammatory monocytes¹⁰⁹, increasing the likelihood that these regulatory cells will be affected by fingolimod. Macrophages that are exposed to fingolimod in culture polarize towards an M2-like phenotype¹¹⁰. Similarly, treatment of murine monocytes with fingolimod increases the recruitment of regulatory monocytes to blood vessels during remodelling in mice¹⁰⁹. In macrophages that are generated from human monocytes,

production of the proinflammatory marker IL-12 is reduced by exposure to fingolimod¹¹¹. In activated microglia, fingolimod downregulates production of proinflammatory cytokines and increases production of BDNF¹¹². Fingolimod also reduces the migratory capacity of dendritic cells, probably by decreasing expression of CCR6 (REF. 113). Furthermore, evidence suggests that fingolimod alters myeloid cells to become not only regulatory, but also more reparative: in a tissue culture model of demyelination, fingolimod modulated microglial activation in a way that boosted markers of remyelination¹¹⁴.

In EAE mice, fingolimod treatment results in accumulation of monocytes in the spleen and bone marrow, suggesting that it restricts egress of monocytes¹¹⁵. In activated monocytes in EAE, fingolimod reduces expression of CD40 and production of TNF¹¹⁵.

Fingolimod has also been shown to affect myeloid cells in patients with MS. In one study, treatment with fingolimod did not reduce the absolute numbers of CD11c⁺ dendritic cells and monocytes, but did diminish the capacity of these cells to produce proinflammatory cytokines¹¹⁶. Furthermore, monocytes cultured from fingolimod-treated patients with MS exhibited low expression of the proinflammatory microRNA miR-155 (REF. 117).

Dimethylfumarate

Dimethylfumarate is absorbed through the small intestine and metabolized rapidly to its active metabolite, monomethylfumarate. Besides reducing generation of proinflammatory T cells, dimethylfumarate activates the Nrf-2 antioxidant pathway, thereby protecting neural cells from oxidative damage¹¹⁸.

Evidence that dimethylfumarate affects myeloid cells is currently limited, but it does exist. For example, dimethylfumarate has been shown to inhibit the maturation of dendritic cells and decrease expression of MHC and co-stimulatory molecules in these cells; as a result, fewer T_H1 and T_H17 cells are generated during antigen presentation¹¹⁹. Similarly, dimethylfumarate seems to inhibit production of nitric oxide and proinflammatory cytokines by microglia¹²⁰, and to suppress the release of neurotoxins from HIV-infected macrophages¹²¹; overall, these responses result in less-toxic myeloid cells. In EAE, dimethylfumarate reduces the severity of disease, an effect that is associated with a reduction in macrophage infiltration into the spinal cord¹²².

In patients with MS who were receiving dimethylfumarate therapy, 12 months of treatment was associated with a reduction in lymphocyte counts, although monocyte numbers were not altered¹²³. Increased numbers and/or activity of regulatory dendritic cells has been documented in MS patients on dimethylfumarate treatment¹²⁴. A study published in 2016 showed that monocytes from dimethylfumarate-treated MS patients expressed miR-155 at lower levels than did monocytes from untreated patients¹¹⁷. In the same study, dimethylfumarate treatment of lipopolysaccharide-activated monocytes or microglia in culture restricted

the expected increase in TNF and IL-6 in activated cells, demonstrating its direct inhibitory effects on myeloid cells.

Teriflunomide

Teriflunomide is an inhibitor of the mitochondrial enzyme dihydroorotate dehydrogenase (DHODH)¹²⁵. Fast-proliferating lymphocytes are dependent on DHODH and are, therefore, targets of this drug¹²⁵.

In EAE, teriflunomide reduces the numbers of T cells and monocytes that infiltrate the CNS, and improves disease outcomes^{126,127}. In culture, teriflunomide increases the secretion of IL-10 by microglia¹²⁶ and inhibits microglia-driven proliferation of T cells¹²⁶. Teriflunomide also inhibits IL-6 and CCL2 production by activated human monocytes¹²⁸. To date, little is known about the effects of teriflunomide therapy on myeloid cells in patients with MS.

Laquinimod

Laquinimod is an immunomodulator that diffuses across the blood–brain barrier, and is an emerging MS therapy¹²⁴. In tissue culture, laquinimod decreases the levels of proinflammatory cytokines and MMP-9 in monocytes and microglia^{32,129,130}. In addition, laquinimod reduces the capacity of dendritic cells to induce CD4⁺ T-cell activation¹³¹, decreases chemokine production by dendritic cells¹³¹, and suppresses the migration of monocytes in tissue culture studies¹³¹.

In mice, laquinimod modulates the proinflammatory activity and migratory capacity of myeloid cells^{32,129,130}, and adoptive transfer of laquinimod-exposed monocytes, which have a regulatory phenotype, ameliorates EAE¹³². Several studies have shown that laquinimod inhibits activation of microglia in culture and in EAE^{124,129}.

In a study of patients with MS, serum levels of BDNF were higher after laquinimod treatment than at baseline or in placebo-treated patients; the elevated BDNF level was thought to result from the induction of a regulatory myeloid phenotype by laquinimod¹³². However, another study showed that laquinimod treatment of patients with MS did not alter the abundance of monocyte subtypes and/or the functional activity of monocytes¹³³.

Monoclonal antibodies

Several monoclonal antibodies, most of which are targeted to T cells or B cells, are used as therapies in MS. Some evidence suggests that these antibodies also affect myeloid cells. In particular, natalizumab binds to α 4-integrin not only on lymphocytes, but also on myeloid cells, thereby blocking their migration into the CNS¹³⁴. Patients with MS who receive natalizumab treatment for 24 months have high numbers of circulating monocytes and lymphocytes, probably because these cells do not enter the CNS¹³⁵. Whether the increased levels of monocytes are accompanied by a bias towards the regulatory phenotype remains unknown.

Possible effects of other monoclonal antibodies on myeloid cells are less clear. Alemtuzumab is a humanized monoclonal antibody that targets CD52, a cell

surface glycoprotein expressed by many leukocyte subsets¹³⁶. Alemtuzumab depletes lymphocytes through complement-mediated lysis of CD52⁺ cells. Infusion of a single dose of alemtuzumab causes long-lasting depletion of lymphocytes, but not of monocytes¹³⁷; whether alemtuzumab profoundly affects the activity of myeloid cells remains unknown.

Targeting of B cells with monoclonal antibodies in MS is an active area of research, as B cells are now known to be important in the pathogenesis of this condition¹³⁸. A study published in 2015 identified a subset of GM-CSF-expressing memory B cells that were more abundant in patients with MS than in healthy controls, and that activated myeloid cells in culture¹³⁹. Treatment of these patients with rituximab, a monoclonal antibody that depletes B cells by targeting CD20 (REF. 140), reduced GM-CSF-dependent proinflammatory myeloid responses. Thus, by depleting B cells, rituximab treatment has the downstream consequence of controlling myeloid cell activity. Nevertheless, much remains unknown about the effects of rituximab, or the humanized anti-CD20 monoclonal antibody ocrelizumab, on myeloid cells.

Glucocorticoids

Glucocorticoids, including methylprednisolone, are commonly used to treat acute relapses of MS. These agents act on glucocorticoid receptors, which are expressed on many cell types, so their effects on cellular processes are widespread and diverse. Glucocorticoids affect various subsets of immune cells, the transcriptional activity of immune-related genes, and many immune pathways (reviewed elsewhere¹⁴¹), with effects that include reduction of the proliferative rate of leukocytes, apoptosis of leukocytes, suppression of proinflammatory cytokine production by T_H1 and T_H17 cells, and improvement of the integrity of the blood-brain barrier owing to reductions in protease activity and leukocyte trafficking.

The effects of glucocorticoids also extend to myeloid cells. It has been known for over 30 years that glucocorticoids reduce the numbers of circulating monocytes and suppress multiple functions of activated myeloid cells, including their accumulation at sites of inflammation and their release of cytokines^{142,143}. The capacity of glucocorticoids to regulate proinflammatory versus regulatory cytokines might differ: one report indicates that macrophages exposed to dexamethasone inhibit the production of proinflammatory IL-12 while increasing that of the regulatory cytokine IL-4 in co-cultured T cells¹⁴⁴.

In mice, treatment with liposome-encapsulated glucocorticoids ameliorated EAE, and this effect was associated with repression of proinflammatory macrophage activity and increased expression of genes associated with the regulatory 'M2' phenotype¹⁴⁵. Another study showed that a low dose of polyethylene glycosylated methylprednisolone reduced the severity of EAE to a similar extent to high-dose free methylprednisolone, an effect that was accompanied by a significant reduction in macrophage and microglial accumulation in

the CNS¹⁴⁶. Thus, the therapeutic effect of glucocorticoids in EAE mice is associated with a decrease in the numbers of myeloid cells in the CNS, or enrichment of regulatory subpopulations.

Studies have also shown that glucocorticoids have a beneficial effect on myeloid cells in patients with MS. In one study, treatment with high doses of glucocorticoids not only reduced T cell activity, but also reduced the numbers of CD14⁺ circulating monocytes, thereby ameliorating their proinflammatory effects¹⁴⁷. Another study showed that brief treatment with intravenous methylprednisolone at the time of MS relapse led to a significant decline in numbers of circulating plasmacytoid dendritic cells, and a trend towards decreased numbers of circulating myeloid dendritic cells¹⁴⁸; these findings suggest that methylprednisolone acts — at least in part — on myeloid cells to treat MS relapses. Another study showed that treatment with 1 g of intravenous methylprednisolone for 4 days at the time of an acute MS relapse increased levels of transcripts encoding the regulatory cytokine IL-10 in peripheral blood mononuclear cells, and decreased levels of TNF and IFN- γ transcripts¹⁴⁹. Serum levels of IL-10 were also increased by glucocorticoid treatment.

Conclusions

Activation of immune cells and their infiltration into the CNS are prominent features of MS pathogenesis. T cells and B cells have received considerable attention in this context, but it is now clear that myeloid cells also have vital roles in the immunopathology of MS. Indeed, myeloid cells could be the direct or indirect targets of several immunomodulators that are used for the treatment of MS. Myeloid cells have probably received less attention in this context because their numbers are replenished quickly, in contrast to the long-lasting lymphocyte depletion that results from some MS medications.

We have grouped monocytes, macrophages, dendritic cells and microglia together under the single umbrella of myeloid cells, but these cell types could be differentially affected by immunomodulators, and this possibility requires further examination. Another important consideration is that although excessively activated myeloid cells contribute to MS pathology, aspects of myeloid cell function are beneficial; such aspects include their production of neurotrophic factors, their removal of myelin and cellular debris¹⁵⁰, and their protection of neurons⁶⁰. Therapeutic interventions that block the damaging effects of myeloid cells in disease while preserving their beneficial functions would be a valuable advance in the management of MS. Some of the medications that we have discussed have such properties; for example, glatiramer acetate reduces proinflammatory myeloid properties by generating regulatory myeloid cells⁴². However, further therapeutics that are targeted specifically to myeloid cells would be an advantage. A greater understanding of the complex properties of myeloid cells in health and in MS will aid the development of such therapies and improve the management and prognosis of MS.

1. Lucchinetti, C. *et al.* Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann. Neurol.* **47**, 707–717 (2000).
2. Prineas, J. W. *et al.* Immunopathology of secondary-progressive multiple sclerosis. *Ann. Neurol.* **50**, 646–657 (2001).
3. Howell, O. W. *et al.* Activated microglia mediate axoglial disruption that contributes to axonal injury in multiple sclerosis. *J. Neuropathol. Exp. Neurol.* **69**, 1017–1033 (2010).
4. Strachan-Whaley, M., Rivest, S. & Yong, V. W. Interactions between microglia and T cells in multiple sclerosis pathobiology. *J. Interferon Cytokine Res.* **34**, 615–622 (2014).
5. Geissmann, F. *et al.* Development of monocytes, macrophages, and dendritic cells. *Science* **327**, 656–661 (2010).
- An excellent review on the types and origins of myeloid cells.**
6. Ginhoux, F. *et al.* Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **330**, 841–845 (2010).
- Seminal work that revealed the origin and development of microglia.**
7. Prinz, M., Tay, T. L., Wolf, Y. & Jung, S. Microglia: unique and common features with other tissue macrophages. *Acta Neuropathol.* **128**, 319–331 (2014).
8. Kierdorf, K. *et al.* Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat. Neurosci.* **16**, 273–280 (2013).
- A pivotal paper that describes the transcription factors that regulate the early development of microglia.**
9. Schulz, C. *et al.* A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science* **336**, 86–90 (2012).
- This paper showed for the first time that microglia precursor cells are Myb-independent, differentiating them from cells that are generated from haematopoietic stem cells.**
10. Gomez Perdiguero, E. *et al.* Tissue-resident macrophages originate from yolk-sac-derived erythromyeloid progenitors. *Nature* **518**, 547–551 (2015).
11. Prinz, M., Priller, J., Sisodia, S. S. & Ransohoff, R. M. Heterogeneity of CNS myeloid cells and their roles in neurodegeneration. *Nat. Neurosci.* **14**, 1227–1235 (2011).
12. Gordon, S. & Taylor, P. R. Monocyte and macrophage heterogeneity. *Nat. Rev. Immunol.* **5**, 953–964 (2005).
13. Delneste, Y. *et al.* Interferon- γ switches monocyte differentiation from dendritic cells to macrophages. *Blood* **101**, 143–150 (2003).
14. Sorokin, L. The impact of the extracellular matrix on inflammation. *Nat. Rev. Immunol.* **10**, 712–723 (2010).
15. Agrawal, S. M. *et al.* Extracellular matrix metalloproteinase inducer shows active perivascular cuffs in multiple sclerosis. *Brain* **136**, 1760–1777 (2013).
16. Tran, E. H., Hoekstra, K., van Rooijen, N., Dijkstra, C. D. & Owens, T. Immune invasion of the central nervous system parenchyma and experimental allergic encephalomyelitis, but not leukocyte extravasation from blood, are prevented in macrophage-depleted mice. *J. Immunol.* **161**, 3767–3775 (1998).
17. Gordon, S. & Martinez, F. O. Alternative activation of macrophages: mechanism and functions. *Immunity* **32**, 593–604 (2010).
18. Murray, P. J. *et al.* Macrophage activation and polarization: nomenclature and experimental guidelines. *Immunity* **41**, 14–20 (2014).
- This review calls for consensus in the nomenclature to describe the various activation states of macrophages.**
19. Xue, J. *et al.* Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity* **40**, 274–288 (2014).
- Through RNA profiling in macrophages exposed to different activators, this paper reveals a broad spectrum of activated macrophages.**
20. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep.* **6**, 13 (2014).
21. Agrawal, S. M., Silva, C., Tourtellotte, W. W. & Yong, V. W. EMMPRIN: a novel regulator of leukocyte transmigration into the CNS in multiple sclerosis and experimental autoimmune encephalomyelitis. *J. Neurosci.* **31**, 669–677 (2011).
22. Nimmerjahn, A., Kirchhoff, F. & Helmchen, F. Resting microglial cells are highly dynamic surveillants of brain parenchyma *in vivo*. *Science* **308**, 1314–1318 (2005).
- A seminal paper in which real-time videomicroscopy was used to show for the first time that microglial processes constantly move and survey their environment *in vivo*.**
23. Czeh, M., Gressens, P. & Kaindl, A. M. The yin and yang of microglia. *Dev. Neurosci.* **33**, 199–209 (2011).
24. Saijo, K. & Glass, C. K. Microglial cell origin and phenotypes in health and disease. *Nat. Rev. Immunol.* **11**, 775–787 (2011).
25. Rawji, K. S. & Yong, V. W. The benefits and detriments of macrophages/microglia in models of multiple sclerosis. *Clin. Dev. Immunol.* **2013**, 948976 (2013).
26. Casano, A. M. & Peri, F. Microglia: multitasking specialists of the brain. *Dev. Cell* **32**, 469–477 (2015).
27. Reizis, B., Bunin, A., Ghosh, H. S., Lewis, K. L. & Sisirak, V. Plasmacytoid dendritic cells: recent progress and open questions. *Annu. Rev. Immunol.* **29**, 163–183 (2011).
28. Miller, S. D., McMahon, E. J., Schreiner, B. & Bailey, S. L. Antigen presentation in the CNS by myeloid dendritic cells drives progression of relapsing experimental autoimmune encephalomyelitis. *Ann. N. Y. Acad. Sci.* **1103**, 179–191 (2007).
29. Duraes, F. V. *et al.* pDC therapy induces recovery from EAE by recruiting endogenous pDC to sites of CNS inflammation. *J. Autoimmun.* **67**, 8–18 (2016).
30. King, I. L., Dickendesher, T. L. & Segal, B. M. Circulating Ly-6C⁺ myeloid precursors migrate to the CNS and play a pathogenic role during autoimmune demyelinating disease. *Blood* **113**, 3190–3197 (2009).
31. Mildner, A. *et al.* CCR2⁺ Ly-6C^{hi} monocytes are crucial for the effector phase of autoimmunity in the central nervous system. *Brain* **132**, 2487–2500 (2009).
32. Mishra, M. K., Wang, J., Silva, C., Mack, M. & Yong, V. W. Kinetics of proinflammatory monocytes in a model of multiple sclerosis and its perturbation by laquinimod. *Am. J. Pathol.* **181**, 642–651 (2012).
33. Ajami, B., Bennett, J. L., Krieger, C., McNagny, K. M. & Rossi, F. M. Infiltrating monocytes trigger EAE progression, but do not contribute to the resident microglia pool. *Nat. Neurosci.* **14**, 1142–1149 (2011).
- A landmark paper showing that the microglial population is not replenished from peripheral immune cells in adulthood.**
34. Agrawal, S. *et al.* Dystroglycan is selectively cleaved at the parenchymal basement membrane at sites of leukocyte extravasation in experimental autoimmune encephalomyelitis. *J. Exp. Med.* **203**, 1007–1019 (2006).
35. Yong, V. W., Power, C., Forsyth, P. & Edwards, D. R. Metalloproteinases in biology and pathology of the nervous system. *Nat. Rev. Neurosci.* **2**, 502–511 (2001).
36. Nuttall, R. K. *et al.* Metalloproteinases are enriched in microglia compared with leukocytes and they regulate cytokine levels in activated microglia. *Glia* **55**, 516–526 (2007).
37. Brosnan, C. F., Bornstein, M. B. & Bloom, B. R. The effects of macrophage depletion on the clinical and pathologic expression of experimental allergic encephalomyelitis. *J. Immunol.* **126**, 614–620 (1981).
38. Huitinga, I., van Rooijen, N., de Groot, C. J., Uitendaele, B. M. & Dijkstra, C. D. Suppression of experimental allergic encephalomyelitis in Lewis rats after elimination of macrophages. *J. Exp. Med.* **172**, 1025–1033 (1990).
39. Bauer, J. *et al.* The role of macrophages, perivascular cells, and microglial cells in the pathogenesis of experimental autoimmune encephalomyelitis. *Glia* **15**, 437–446 (1995).
40. Sloka, S., Metz, L. M., Hader, W., Starreveld, Y. & Yong, V. W. Reduction of microglial activity in a model of multiple sclerosis by dipyrindamole. *J. Neuroinflamm.* **10**, 89 (2013).
41. Mikita, J. *et al.* Altered M1/M2 activation patterns of monocytes in severe relapsing experimental rat model of multiple sclerosis. Amelioration of clinical status by M2 activated monocyte administration. *Mult. Scler.* **17**, 2–15 (2011).
42. Weber, M. S. *et al.* Type II monocytes modulate T cell-mediated central nervous system autoimmune disease. *Nat. Med.* **13**, 935–943 (2007).
- This manuscript highlights the generation of regulatory myeloid cells by glatiramer acetate treatment, leading to the production of T_H2 cells commonly associated with this medication.**
43. Benveniste, E. N. Role of macrophages/microglia in multiple sclerosis and experimental allergic encephalomyelitis. *J. Mol. Med. (Berl.)* **75**, 165–173 (1997).
44. Sosa, R. A., Murphey, C., Robinson, R. R. & Forsthuber, T. G. IFN- γ ameliorates autoimmune encephalomyelitis by limiting myelin lipid peroxidation. *Proc. Natl. Acad. Sci. USA* **112**, E5038–E5047 (2015).
45. Nikic, I. *et al.* A reversible form of axon damage in experimental autoimmune encephalomyelitis and multiple sclerosis. *Nat. Med.* **17**, 495–499 (2011).
46. Mossakowski, A. A. *et al.* Tracking CNS and systemic sources of oxidative stress during the course of chronic neuroinflammation. *Acta Neuropathol.* **130**, 799–814 (2015).
47. van Horssen, J., Witte, M. E., Schreibeit, G. & de Vries, H. E. Radical changes in multiple sclerosis pathogenesis. *Biochim. Biophys. Acta* **1812**, 141–150 (2011).
48. Yamasaki, R. *et al.* Differential roles of microglia and monocytes in the inflamed central nervous system. *J. Exp. Med.* **211**, 1533–1549 (2014).
- This important paper revealed the different roles of microglia and macrophages in the experimental autoimmune encephalitis brain.**
49. Brosnan, C. F., Sacks, H. J., Goldschmidt, R. C., Goldmuntz, E. A. & Norton, W. T. Prazosin treatment during the effector stage of disease suppresses experimental autoimmune encephalomyelitis in the Lewis rat. *J. Immunol.* **137**, 3451–3456 (1986).
50. Shaked, I. *et al.* Transcription factor Nr4a1 couples sympathetic and inflammatory cues in CNS-recruited macrophages to limit neuroinflammation. *Nat. Immunol.* **16**, 1228–1234 (2015).
- An important paper that shows that noradrenaline from macrophages helps promote the migration of leukocytes into the CNS, thereby linking stress and neuroinflammation.**
51. Ponomarev, E. D., Shriver, L. P., Maresz, K. & Dittel, B. N. Microglial cell activation and proliferation precedes the onset of CNS autoimmunity. *J. Neurosci. Res.* **81**, 374–389 (2005).
52. Rasmussen, S. *et al.* Persistent activation of microglia is associated with neuronal dysfunction of callosal projecting pathways and multiple sclerosis-like lesions in relapsing–remitting experimental autoimmune encephalomyelitis. *Brain* **130**, 2816–2829 (2007).
53. Takeuchi, H. *et al.* Tumor necrosis factor- α induces neurotoxicity via glutamate release from hemichannels of activated microglia in an autocrine manner. *J. Biol. Chem.* **281**, 21362–21368 (2006).
54. Aloisi, F., Ria, F., Penna, G. & Adorini, L. Microglia are more efficient than astrocytes in antigen processing and in Th1 but not Th2 cell activation. *J. Immunol.* **160**, 4671–4680 (1998).
55. Aloisi, F. *et al.* Relative efficiency of microglia, astrocytes, dendritic cells and B cells in naive CD4⁺ T cell priming and Th1/Th2 cell restimulation. *Eur. J. Immunol.* **29**, 2705–2714 (1999).
56. Davalos, D. *et al.* Fibrinogen-induced perivascular microglial clustering is required for the development of axonal damage in neuroinflammation. *Nat. Commun.* **3**, 1227 (2012).
57. Heppner, F. L. *et al.* Experimental autoimmune encephalomyelitis repressed by microglial paralysis. *Nat. Med.* **11**, 146–152 (2005).
58. Goldmann, T. *et al.* A new type of microglia gene targeting shows TAK1 to be pivotal in CNS autoimmune inflammation. *Nat. Neurosci.* **16**, 1618–1626 (2013).
59. Neumann, J. *et al.* Microglia cells protect neurons by direct engulfment of invading neutrophil granulocytes: a new mechanism of CNS immune privilege. *J. Neurosci.* **28**, 5965–5975 (2008).
60. Chen, Z. *et al.* Microglial displacement of inhibitory synapses provides neuroprotection in the adult brain. *Nat. Commun.* **5**, 4486 (2014).
61. Ferguson, B., Matyszak, M. K., Esiri, M. M. & Perry, V. H. Axonal damage in acute multiple sclerosis lesions. *Brain* **120**, 393–399 (1997).
62. Politis, M. *et al.* Increased PK11195 PET binding in the cortex of patients with MS correlates with disability. *Neurology* **79**, 523–530 (2012).
- An important paper that describes the detection of activated microglia and macrophages through PET imaging in people with multiple sclerosis.**

63. Kutzelnigg, A. *et al.* Cortical demyelination and diffuse white matter injury in multiple sclerosis. *Brain* **128**, 2705–2712 (2005).
64. van Noort, J. M. *et al.* Preactive multiple sclerosis lesions offer novel clues for neuroprotective therapeutic strategies. *CNS Neurol. Disord. Drug Targets* **10**, 68–81 (2011).
65. Vogel, D. Y. *et al.* Macrophages in inflammatory multiple sclerosis lesions have an intermediate activation status. *J. Neuroinflamm.* **10**, 35 (2013).
66. Kooi, E. J., Srijbis, E. M., van der Valk, P. & Geurts, J. J. Heterogeneity of cortical lesions in multiple sclerosis: clinical and pathologic implications. *Neurology* **79**, 1369–1376 (2012).
67. Singh, S. *et al.* Microglial nodules in early multiple sclerosis white matter are associated with degenerating axons. *Acta Neuropathol.* **125**, 595–608 (2013).
68. Peferoen, L. A. *et al.* Activation status of human microglia is dependent on lesion formation stage and remyelination in multiple sclerosis. *J. Neuropathol. Exp. Neurol.* **74**, 48–63 (2015).
69. Kouwenvloot, M., Teleshova, N., Ozenci, V., Press, R. & Link, H. Monocytes in multiple sclerosis: phenotype and cytokine profile. *J. Neuroimmunol.* **112**, 197–205 (2001).
70. Huang, Y. M. *et al.* Multiple sclerosis is associated with high levels of circulating dendritic cells secreting pro-inflammatory cytokines. *J. Neuroimmunol.* **99**, 82–90 (1999).
71. Makhlof, K., Weiner, H. L. & Khoury, S. J. Increased percentage of IL-12⁺ monocytes in the blood correlates with the presence of active MRI lesions in MS. *J. Neuroimmunol.* **119**, 145–149 (2001).
72. Waschbisch, A. *et al.* Pivotal role for CD16⁺ monocytes in immune surveillance of the central nervous system. *J. Immunol.* **196**, 1558–1567 (2016).
73. Kivisakk, P. *et al.* Localizing central nervous system immune surveillance: meningeal antigen-presenting cells activate T cells during experimental autoimmune encephalomyelitis. *Ann. Neurol.* **65**, 457–469 (2009).
74. Croxford, A. L. *et al.* The cytokine GM-CSF drives the inflammatory signature of CCR2⁺ monocytes and licenses autoimmunity. *Immunity* **43**, 502–514 (2015). **A key paper that links T_H17 cells and the production of GM-CSF with the downstream generation of proinflammatory monocytes that mediate CNS inflammation.**
75. Codarri, L., Greter, M. & Becher, B. Communication between pathogenic T cells and myeloid cells in neuroinflammatory disease. *Trends Immunol.* **34**, 114–119 (2013).
76. Ponomarev, E. D. *et al.* GM-CSF production by autoreactive T cells is required for the activation of microglial cells and the onset of experimental autoimmune encephalomyelitis. *J. Immunol.* **178**, 39–48 (2007).
77. Codarri, L. *et al.* RORγt drives production of the cytokine GM-CSF in helper T cells, which is essential for the effector phase of autoimmune neuroinflammation. *Nat. Immunol.* **12**, 560–567 (2011).
78. El-Behi, M. *et al.* The encephalitogenicity of T_H17 cells is dependent on IL-1 α and IL-23-induced production of the cytokine GM-CSF. *Nat. Immunol.* **12**, 568–575 (2011).
79. Croxford, A. L., Spath, S. & Becher, B. GM-CSF in neuroinflammation: licensing myeloid cells for tissue damage. *Trends Immunol.* **36**, 651–662 (2015). **An excellent review that discusses the critical role of myeloid cells in tissue damage in the CNS.**
80. Bruck, W. *et al.* Therapeutic decisions in multiple sclerosis: moving beyond efficacy. *JAMA Neurol.* **70**, 1315–1324 (2013).
81. Stuve, O. *et al.* Pharmacological treatment of early multiple sclerosis. *Drugs* **68**, 73–83 (2008).
82. Biber, K., Moller, T., Boddeke, E. & Prinz, M. Central nervous system myeloid cells as drug targets: current status and translational challenges. *Nat. Rev. Drug Discov.* **15**, 110–124 (2016). **This comprehensive review discusses the effector functions of myeloid cells and the approaches to targeting the proinflammatory activity of myeloid cells.**
83. Dhib-Jalbut, S. & Marks, S. Interferon-beta mechanisms of action in multiple sclerosis. *Neurology* **74**, S17–S24 (2010).
84. Markmann, S. *et al.* Interferon- β up-regulates the expression of co-stimulatory molecules CD80, CD86 and CD40 on monocytes: significance for treatment of multiple sclerosis. *Clin. Exp. Immunol.* **138**, 499–506 (2004).
85. Liu, Z., Pelfrey, C. M., Coteleur, A., Lee, J. C. & Rudick, R. A. Immunomodulatory effects of interferon beta-1a in multiple sclerosis. *J. Neuroimmunol.* **112**, 153–162 (2001).
86. Ramgolam, V. S., Sha, Y., Jin, J., Zhang, X. & Markovic-Plese, S. IFN- β inhibits human Th17 cell differentiation. *J. Immunol.* **183**, 5418–5427 (2009).
87. Yen, J. H., Kong, W. & Ganea, D. IFN- β inhibits dendritic cell migration through STAT1-mediated transcriptional suppression of CCR7 and matrix metalloproteinase 9. *J. Immunol.* **184**, 3478–3486 (2010). **An important paper that shows the direct effect of IFN- β on dendritic cells.**
88. Galboiz, Y., Shapiro, S., Lahat, N. & Miller, A. Modulation of monocytes matrix metalloproteinase-2, MT1-MMP and TIMP-2 by interferon- γ and - β : implications to multiple sclerosis. *J. Neuroimmunol.* **131**, 191–200 (2002).
89. Schreiner, B. *et al.* Interferon- β enhances monocyte and dendritic cell expression of B7-H1 (PD-L1), a strong inhibitor of autologous T-cell activation: relevance for the immune modulatory effect in multiple sclerosis. *J. Neuroimmunol.* **155**, 172–182 (2004).
90. Comabella, M. *et al.* A type I interferon signature in monocytes is associated with poor response to interferon- β in multiple sclerosis. *Brain* **132**, 3353–3365 (2009).
91. Yen, J. H. & Ganea, D. Interferon β induces mature dendritic cell apoptosis through caspase-11/caspase-3 activation. *Blood* **114**, 1344–1354 (2009).
92. Guo, B., Chang, E. Y. & Cheng, G. The type I IFN induction pathway constrains Th17-mediated autoimmune inflammation in mice. *J. Clin. Invest.* **118**, 1680–1690 (2008).
93. Prinz, M. *et al.* Distinct and nonredundant *in vivo* functions of IFNAR on myeloid cells limit autoimmunity in the central nervous system. *Immunity* **28**, 675–686 (2008). **A key paper that implicates myeloid cells as important cellular targets of IFN- β signalling to ameliorate neuroinflammation.**
94. Hussien, Y., Sanna, A., Soderstrom, M., Link, H. & Huang, Y. M. Multiple sclerosis: expression of CD1a and production of IL-12p70 and IFN- γ by blood mononuclear cells in patients on combination therapy with IFN- β and glatiramer acetate compared to monotherapy with IFN- β . *Mult. Scler.* **10**, 16–25 (2004).
95. Lucas, M. *et al.* Regulation by interferon β -1a of reactive oxygen metabolites production by lymphocytes and monocytes and serum sulfhydryls in relapsing multiple sclerosis patients. *Neurochem. Int.* **42**, 67–71 (2003).
96. Hamamcioglu, K. & Reder, A. T. Interferon- β regulates cytokines and BDNF: greater effect in relapsing than in progressive multiple sclerosis. *Mult. Scler.* **13**, 459–470 (2007).
97. Waschbisch, A. *et al.* Interferon beta and vitamin D synergize to induce immunoregulatory receptors on peripheral blood monocytes of multiple sclerosis patients. *PLoS ONE* **9**, e115488 (2014).
98. Weber, M. S. *et al.* Multiple sclerosis: glatiramer acetate inhibits monocyte reactivity *in vitro* and *in vivo*. *Brain* **127**, 1370–1378 (2004).
99. Kim, H. J. *et al.* Type 2 monocyte and microglia differentiation mediated by glatiramer acetate therapy in patients with multiple sclerosis. *J. Immunol.* **172**, 7144–7153 (2004). **One of the first papers to describe the generation of regulatory myeloid cells by glatiramer acetate.**
100. Sellebjerg, F. *et al.* Dendritic cell, monocyte and T cell activation and response to glatiramer acetate in multiple sclerosis. *Mult. Scler.* **19**, 179–187 (2013).
101. Iarlori, C. *et al.* Reduction of free radicals in multiple sclerosis: effect of glatiramer acetate (Copaxone). *Mult. Scler.* **14**, 739–748 (2008).
102. Pul, R. *et al.* Glatiramer acetate modulates TNF- α and IL-10 secretion in microglia and promotes their phagocytic activity. *J. Neuroimmune Pharmacol.* **6**, 381–388 (2011).
103. Pul, R. *et al.* Glatiramer acetate increases phagocytic activity of human monocytes *in vitro* and in multiple sclerosis patients. *PLoS ONE* **7**, e1867 (2012).
104. Burger, D. *et al.* Glatiramer acetate increases IL-1 β receptor antagonist but decreases T cell-induced IL-1 β in human monocytes and multiple sclerosis. *Proc. Natl Acad. Sci. USA* **106**, 4355–4359 (2009). **This study shows that human monocytes are responsive to glatiramer acetate and that they increase their regulatory properties as a result.**
105. Ayers, C. L. *et al.* Modulation of immune function occurs within hours of therapy initiation for multiple sclerosis. *Clin. Immunol.* **147**, 105–119 (2013).
106. Ratcliff, J. N. *et al.* Decreased microglial activation in MS patients treated with glatiramer acetate. *J. Neurol.* **259**, 1199–1205 (2012).
107. Chun, J. & Hartung, H. P. Mechanism of action of oral fingolimod (FTY720) in multiple sclerosis. *Clin. Neuropharmacol.* **33**, 91–101 (2010).
108. Muller, H. *et al.* The immunomodulator FTY720 interferes with effector functions of human monocyte-derived dendritic cells. *Eur. J. Immunol.* **35**, 533–545 (2005).
109. Awojoodu, A. O. *et al.* Sphingosine 1-phosphate receptor 3 regulates recruitment of anti-inflammatory monocytes to microvessels during implant arteriogenesis. *Proc. Natl Acad. Sci. USA* **110**, 13785–13790 (2013).
110. Hughes, J. E. *et al.* Sphingosine-1-phosphate induces an antiinflammatory phenotype in macrophages. *Circ. Res.* **102**, 950–958 (2008).
111. Durafourt, B. A. *et al.* Differential responses of human microglia and blood-derived myeloid cells to FTY720. *J. Neuroimmunol.* **230**, 10–16 (2011).
112. Noda, H., Takeuchi, H., Mizuno, T. & Suzumura, A. Fingolimod phosphate promotes the neuroprotective effects of microglia. *J. Neuroimmunol.* **256**, 13–18 (2013). **The study that demonstrated that microglia are altered by fingolimod exposure to generate growth factors.**
113. Al-Jaderi, Z. & Maghazachi, A. A. Effects of vitamin D₃, calcipotriol and FTY720 on the expression of surface molecules and cytolytic activities of human natural killer cells and dendritic cells. *Toxins (Basel)* **5**, 1932–1947 (2013).
114. Jackson, S. J., Giovannoni, G. & Baker, D. Fingolimod modulates microglial activation to augment markers of remyelination. *J. Neuroinflamm.* **8**, 76 (2011). **This study demonstrated that microglia are affected by fingolimod, and that they become pro-reparative as a result.**
115. Lewis, N. D. *et al.* Circulating monocytes are reduced by sphingosine-1-phosphate receptor modulators independently of S1P₅. *J. Immunol.* **190**, 3533–3540 (2013).
116. Luessi, F. *et al.* FTY720 (fingolimod) treatment tips the balance towards less immunogenic antigen-presenting cells in patients with multiple sclerosis. *Mult. Scler.* **21**, 1811–1822 (2015).
117. Michell-Robinson, M. A. *et al.* Effects of fumarates on circulating and CNS myeloid cells in multiple sclerosis. *Ann. Clin. Transl. Neurol.* **3**, 27–41 (2016). **This study showed that monocytes from patients with multiple sclerosis who were treated with dimethylfumarate had reduced expression of the proinflammatory microRNA miR-155.**
118. Linker, R. A. *et al.* Fumaric acid esters exert neuroprotective effects in neuroinflammation via activation of the Nrf2 antioxidant pathway. *Brain* **134**, 678–692 (2011).
119. Peng, H. *et al.* Dimethyl fumarate inhibits dendritic cell maturation via nuclear factor κ B (NF- κ B) and extracellular signal-regulated kinase 1 and 2 (ERK1/2) and mitogen stress-activated kinase 1 (MSK1) signaling. *J. Biol. Chem.* **287**, 28017–28026 (2012).
120. Wilms, H. *et al.* Dimethylfumarate inhibits microglial and astrocytic inflammation by suppressing the synthesis of nitric oxide, IL-1 β , TNF- α and IL-6 in an *in-vitro* model of brain inflammation. *J. Neuroinflamm.* **7**, 30 (2010).
121. Cross, S. A. *et al.* Dimethyl fumarate, an immune modulator and inducer of the antioxidant response, suppresses HIV replication and macrophage-mediated neurotoxicity: a novel candidate for HIV neuroprotection. *J. Immunol.* **187**, 5015–5025 (2011).
122. Schilling, S., Goelz, S., Linker, R., Luehder, F. & Gold, R. Fumaric acid esters are effective in chronic experimental autoimmune encephalomyelitis and suppress macrophage infiltration. *Clin. Exp. Immunol.* **145**, 101–107 (2006).
123. Spencer, C. M., Crabtree-Hartman, E. C., Lehmann-Horn, K., Cree, B. A. & Zamvil, S. S. Reduction of CD8⁺ T lymphocytes in multiple sclerosis patients treated with dimethyl fumarate. *Neurol. Neuroimmunol. Neuroinflamm.* **2**, e76 (2015).
124. Ghoreschi, K. *et al.* Fumarates improve psoriasis and multiple sclerosis by inducing type II dendritic cells. *J. Exp. Med.* **208**, 2291–2303 (2011). **This manuscript suggested for the first time that dimethylfumarate has an effect on the generation of regulatory dendritic cells.**

125. Tanasescu, R., Evangelou, N. & Constantinescu, C. S. Role of oral teriflunomide in the management of multiple sclerosis. *Neuropsychiatr. Dis. Treat.* **9**, 539–553 (2013).
126. Korn, T., Magnus, T., Toyka, K. & Jung, S. Modulation of effector cell functions in experimental autoimmune encephalomyelitis by leflunomide — mechanisms independent of pyrimidine depletion. *J. Leukoc. Biol.* **76**, 950–960 (2004).
127. Ringheim, G. E. *et al.* Teriflunomide attenuates immunopathological changes in the dark agouti rat model of experimental autoimmune encephalomyelitis. *Front. Neurol.* **4**, 169 (2013).
128. Li, L. *et al.* The effects of teriflunomide on lymphocyte subpopulations in human peripheral blood mononuclear cells *in vitro*. *J. Neuroimmunol.* **265**, 82–90 (2013).
129. Mishra, M. K. *et al.* Laquinimod reduces neuroaxonal injury through inhibiting microglial activation. *Ann. Clin. Transl. Neurol.* **1**, 409–422 (2014).
- A manuscript that describes that microglial activation is normalized by laquinimod.**
130. Schulze-Topphoff, U. *et al.* Laquinimod, a quinoline-3-carboxamide, induces type II myeloid cells that modulate central nervous system autoimmunity. *PLoS ONE* **7**, e33797 (2012).
131. Jolivel, V. *et al.* Modulation of dendritic cell properties by laquinimod as a mechanism for modulating multiple sclerosis. *Brain* **136**, 1048–1066 (2013).
132. Thone, J. *et al.* Modulation of autoimmune demyelination by laquinimod via induction of brain-derived neurotrophic factor. *Am. J. Pathol.* **180**, 267–274 (2012).
133. Lund, B. T. *et al.* Assessment of changes in immune measures of multiple sclerosis patients treated with laquinimod. *J. Neuroimmunol.* **263**, 108–115 (2013).
134. Ali, R., Nicholas, R. S. & Muraro, P. A. Drugs in development for relapsing multiple sclerosis. *Drugs* **73**, 625–650 (2013).
135. Planas, R., Jelcic, I., Schippling, S., Martin, R. & Sospedra, M. Natalizumab treatment perturbs memory- and marginal zone-like B-cell homing in secondary lymphoid organs in multiple sclerosis. *Eur. J. Immunol.* **42**, 790–798 (2012).
136. Jones, J. L. & Coles, A. J. Mode of action and clinical studies with alemtuzumab. *Exp. Neurol.* **262**, 37–43 (2014).
137. Coles, A. J. *et al.* The window of therapeutic opportunity in multiple sclerosis: evidence from monoclonal antibody therapy. *J. Neurol.* **253**, 98–108 (2006).
138. Boster, A., Ankeny, D. P. & Racke, M. K. The potential role of B cell-targeted therapies in multiple sclerosis. *Drugs* **70**, 2343–2356 (2010).
139. Li, R. *et al.* Proinflammatory GM-CSF-producing B cells in multiple sclerosis and B cell depletion therapy. *Sci. Transl. Med.* **7**, 310ra166 (2015).
- An excellent paper that showed that the depletion of B cells in patients with multiple sclerosis led to decreases in proinflammatory myeloid responses.**
140. Kausar, F. *et al.* Ocrelizumab: a step forward in the evolution of B-cell therapy. *Expert. Opin. Biol. Ther.* **9**, 889–895 (2009).
141. Schweingruber, N., Reichardt, S. D., Luhder, F. & Reichardt, H. M. Mechanisms of glucocorticoids in the control of neuroinflammation. *J. Neuroendocrinol.* **24**, 174–182 (2012).
142. Joyce, D. A., Steer, J. H. & Abraham, L. J. Glucocorticoid modulation of human monocyte/macrophage function: control of TNF- α secretion. *Inflamm. Res.* **46**, 447–451 (1997).
143. Parrillo, J. E. & Fauci, A. S. Mechanisms of glucocorticoid action on immune processes. *Annu. Rev. Pharmacol. Toxicol.* **19**, 179–201 (1979).
144. DeKruyff, R. H., Fang, Y. & Umetsu, D. T. Corticosteroids enhance the capacity of macrophages to induce Th2 cytokine synthesis in CD4⁺ lymphocytes by inhibiting IL-12 production. *J. Immunol.* **160**, 2231–2237 (1998).
145. Schweingruber, N. *et al.* Liposomal encapsulation of glucocorticoids alters their mode of action in the treatment of experimental autoimmune encephalomyelitis. *J. Immunol.* **187**, 4310–4318 (2011).
146. Lee, D. H. *et al.* Glutathione PEGylated liposomal methylprednisolone (2B3-201) attenuates CNS inflammation and degeneration in murine myelin oligodendrocyte glycoprotein induced experimental autoimmune encephalomyelitis. *J. Neuroimmunol.* **274**, 96–101 (2014).
147. Frisullo, G. *et al.* Glucocorticoid treatment reduces T-bet and pSTAT1 expression in mononuclear cells from relapsing remitting multiple sclerosis patients. *Clin. Immunol.* **124**, 284–293 (2007).
148. Krystyna, M. S. *et al.* Changes in circulating dendritic cells and B-cells in patients with multiple sclerosis relapse during corticosteroid therapy. *J. Neuroimmunol.* **207**, 107–110 (2009).
149. Gayo, A. *et al.* Glucocorticoids increase IL-10 expression in multiple sclerosis patients with acute relapse. *J. Neuroimmunol.* **85**, 122–130 (1998).
150. Wee Yong, V. Inflammation in neurological disorders: a help or a hindrance? *Neuroscientist* **16**, 408–420 (2010).

Acknowledgements

The authors' research is supported by grants from the Canadian Institutes of Health Research, the Alberta Innovates–Health Solutions CRIO Team programme, and the Multiple Sclerosis Society of Canada.

Author contributions

Both authors contributed equally to all aspects of the manuscript.

Competing interests statement

V.W.Y. acknowledges previous unrestricted operating grant funding from EMD-Serono, Novartis and Teva Pharmaceuticals, and previous and current unrestricted educational grants from Biogen-Idec, EMD-Serono, Genzyme, Novartis, Roche and Teva Neuroscience. V.W.Y. has received honoraria for seminar presentations from Biogen-Idec, Genzyme, Novartis and Teva Neuroscience. M.K.M. declares no competing interests.

Transcription factor Nr4a1 couples sympathetic and inflammatory cues in CNS-recruited macrophages to limit neuroinflammation

Iftach Shaked^{1,8}, Richard N Hanna^{1,8}, Helena Shaked¹, Grzegorz Chodaczek², Heba N Nowyhed¹, George Tweet¹, Robert Tacke¹, Alp Bugra Basat¹, Zbigniew Mikulski², Susan Togher¹, Jacqueline Miller¹, Amy Blatchley¹, Shahram Salek-Ardakani³, Martin Darvas⁴, Minna U Kaikkonen⁵, Graham D Thomas¹, Sonia Lai-Wing-Sun⁶, Ayman Rezk⁶, Amit Bar-Or⁶, Christopher K Glass⁷, Hozefa Bandukwala¹ & Catherine C Hedrick¹

The molecular mechanisms that link the sympathetic stress response and inflammation remain obscure. Here we found that the transcription factor Nr4a1 regulated the production of norepinephrine (NE) in macrophages and thereby limited experimental autoimmune encephalomyelitis (EAE), a mouse model of multiple sclerosis. Lack of Nr4a1 in myeloid cells led to enhanced NE production, accelerated infiltration of leukocytes into the central nervous system (CNS) and disease exacerbation *in vivo*. In contrast, myeloid-specific deletion of tyrosine hydroxylase (TH), the rate-limiting enzyme in catecholamine biosynthesis, protected mice against EAE. Furthermore, we found that Nr4a1 repressed autocrine NE production in macrophages by recruiting the corepressor CoREST to the *Th* promoter. Our data reveal a new role for macrophages in neuroinflammation and identify Nr4a1 as a key regulator of catecholamine production by macrophages.

Multiple sclerosis (MS) is an autoimmune inflammatory disease of the central nervous system (CNS) that has long been associated with stress and the sympathoadrenergic response^{1,2}. The sympathetic hormone norepinephrine (NE), a major mediator of the response to physiological and psychological stressors, has been shown to be involved in the pathology of experimental autoimmune encephalitis (EAE), the rodent model of MS^{3–6}. However, the molecular mechanisms that link stress signaling to neuroinflammation remain unclear.

The Nr4a orphan nuclear receptors Nr4a1 (Nur77), Nr4a2 (Nurr1) and Nr4a3 (Nor1) are encoded by early-immEDIATE response genes rapidly induced by a variety of physiological stimuli⁷. Nr4a1, like other members of the Nr4a family, is involved in early sympathetic stress response in the neuroendocrine system^{8–11}. At the same time, Nr4a1 has an important role in leukocytes, in which it is a central regulator of innate and adaptive immune responses. As such, Nr4a1 is involved in the activation and differentiation of macrophages in atherosclerosis^{12,13} and also controls the differentiation and survival of nonclassical Ly6C[–] patrolling monocytes^{14,15}. Furthermore, Nr4a1 is rapidly induced in T cells following activation of the T cell antigen receptor and reflects the strength of signaling via this receptor¹⁶. Thus, Nr4a1 is involved both in immunity and in the stress response and it might therefore represent a key junction in the crosstalk between

the sympathetic nervous system and immune system, particularly in the context of neuroinflammation.

To test that hypothesis, we used an established model of EAE and found that mice lacking Nr4a1 developed accelerated and exacerbated disease that was accompanied by high concentrations of NE and interleukin 6 (IL-6) and early auto-aggressive infiltration of T cells into the CNS. Mechanistically, we discovered that Nr4a1 inhibited macrophage expression of tyrosine hydroxylase (TH), the rate-limiting enzyme for NE production¹⁷. We found that myeloid-specific deletion of TH protected mice from EAE. Our data demonstrate a major role for the production of NE by macrophages in neuroinflammation and identify an important mechanism for the regulation of NE by Nr4a1.

RESULTS

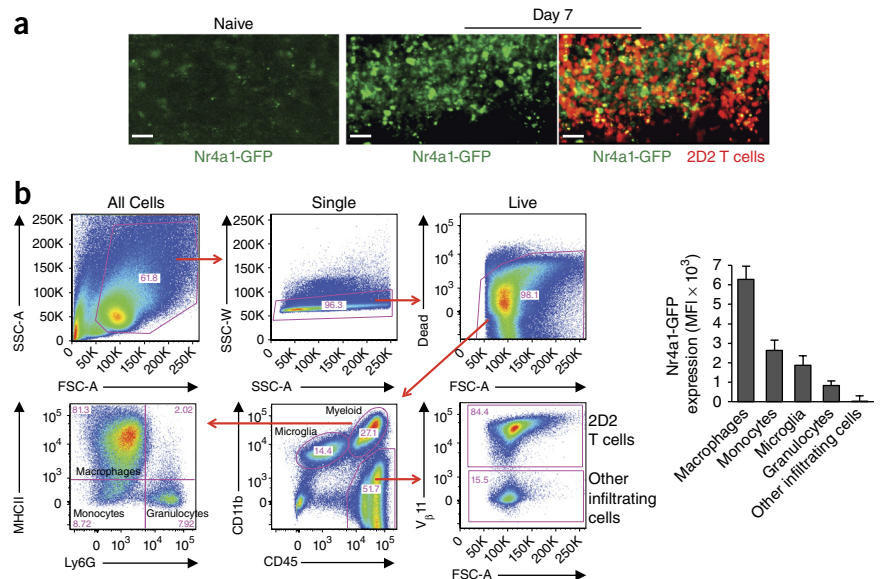
Nr4a1 expression in myeloid cells limits EAE severity

To determine if Nr4a1 has a role in CNS autoimmunity, we analyzed Nr4a1 expression in mice during EAE. We used a passive model of EAE (Supplementary Fig. 1a) that involves the *in vitro* differentiation of autoreactive CD4⁺ T lymphocytes from 2D2 mice (which have transgenic expression of a T cell antigen receptor specific for a myelin oligodendrocyte glycoprotein peptide of amino acids 35–55)¹⁸

¹Division of Inflammation Biology, La Jolla Institute for Allergy and Immunology, La Jolla, California, USA. ²Microscopy Core, La Jolla Institute for Allergy and Immunology, La Jolla, California, USA. ³Department of Pathology, Immunology & Laboratory Medicine, University of Florida College of Medicine, Gainesville, Florida, USA. ⁴Department of Pathology, University of Washington, Seattle, Washington, USA. ⁵Department of Biotechnology and Molecular Medicine, University of Eastern Finland, Kuopio, Finland. ⁶Montreal Neurological Institute, McGill University, Montreal, Canada. ⁷Department of Cellular & Molecular Medicine, University of California San Diego, San Diego, California, USA. ⁸These authors contributed equally to this work. Correspondence should be addressed to I.S. (iftach@lji.org) or C.C.H. (hedrick@lji.org).

Received 24 July; accepted 13 October; published online 2 November 2015; doi:10.1038/ni.3321

Figure 1 High expression of Nr4a1 in myeloid cells in the CNS at the onset of EAE. **(a)** *In vivo* imaging of the induction of Nr4a1 in spinal cord before (naive) or 7 d after (Day 7) the transfer of 2×10^6 2D2-DsRed T cells into Nr4a1-GFP reporter mice. Scale bars, 30 μ m. **(b)** Flow cytometry of leukocytes isolated from the CNS of Nr4a1-GFP reporter mice at day 10 after the induction of EAE: top two rows, gating scheme; below, quantification of Nr4a1-GFP expression, presented as median fluorescence intensity (MFI). Numbers in outlined areas (above) indicate percent cells positive for side scatter (SSC) and forward scatter (FSC) (top left plot); singlets (top middle plot); live cells (top right plot); CD11b⁺CD45⁺ (microglial) cells (top left), CD11b⁺CD45⁺ (myeloid) cells (top right) or CD11b⁺CD45⁺ cells (bottom right), among live cells (bottom middle plot); V β 11⁺ (2D2) cells (top) or V β 11⁻ cells (bottom) (bottom right plot); or MHCII⁺Ly6G⁻ macrophages (top left), MHCII⁺Ly6G⁺ cells (top right), MHCII⁻Ly6G⁺ granulocytes (bottom right) or MHCII⁻Ly6G⁻ monocytes (bottom left) (bottom left plot). Data are representative of three **(a)** or two **(b)** independent experiments (error bars **(b)**, s.e.m. of $n = 6$ mice).



into the T_H1 subset of helper T cells, followed by adoptive transfer of the cells into host mice (Supplementary Fig. 1b). To address Nr4a1 expression during EAE at the cellular level, we induced the disease in mice expressing green fluorescent protein (GFP) under control of the *Nr4a1* promoter (Nr4a1-GFP)¹⁶. By intravital microscopy, we observed significant induction of Nr4a1-GFP expression in the spinal cord at the onset of EAE (Fig. 1a). To determine the identity of the cells expressing Nr4a1-GFP, we used flow cytometry. We identified microglia as CD45^{lo}CD11b⁺ cells¹⁹, whereas the phenotype CD45^{hi}CD11b⁺ identified infiltrating myeloid cells (Fig. 1b). We further gated CD45^{hi}CD11b⁺ infiltrating myeloid populations to distinguish among granulocytes (Ly6G⁺MHCII⁻), monocytes (Ly6G⁻MHCII⁻) and macrophages (Ly6G⁻MHCII⁺) (Fig. 1b). Macrophages identified by flow cytometry in the CNS were also positive for the core tissue macrophage marker CD64 and the receptor MerTK²⁰, while monocytes characteristically expressed some CD64 but not MerTK²¹ (Supplementary Fig. 1c). Nr4a1-GFP had high expression in infiltrating macrophages and, to a lesser extent in infiltrating monocytes, and in resident microglia (Fig. 1b). Nr4a1 expression was relatively low in granulocytes and other infiltrating cells (which were probably non-2D2 lymphocytes) (Fig. 1b).

To address the importance of Nr4a1 expression in CNS autoimmunity, we induced EAE in *Nr4a1*^{-/-} mice²². This resulted in much earlier onset and exacerbated disease development (Fig. 2a and Supplementary Fig. 2a,b), as well as accelerated body mass loss (Supplementary Fig. 2c), compared with that of wild-type mice. *Nr4a1*^{-/-} mice were also more susceptible than wild-type mice to EAE induced by active immunization with myelin oligodendrocyte glycoprotein peptide (amino acids 35–55) (Supplementary Fig. 2d).

Nr4a1 had high expression in myeloid cells in the CNS at the onset of EAE (Fig. 1). Therefore, to address the role of Nr4a1 in myeloid cells, we used mice with specific deletion of loxP-flanked *Nr4a1* alleles (*Nr4a1*^{fl/fl}) by Cre recombinase expressed from either of two well-characterized myeloid-specific transgenes, *LysM*-Cre (*Nr4a1*^{ΔLysM}) or *Csf1r*-Cre (*Nr4a1*^{ΔCsf1r}). We first analyzed the expression of *Nr4a1* mRNA and deletion efficacy in the myeloid cells in these strains. In *Nr4a1*^{fl/fl} mice negative for the Cre-encoding transgenes, *Nr4a1* mRNA had its

highest expression in monocytes and macrophages, with lower expression in granulocytes and microglia (Fig. 2b). Expression of *LysM*-Cre significantly reduced the expression of *Nr4a1* mRNA in macrophages, monocytes and granulocytes but not in microglia, while expression of *Csf1r*-Cre effectively reduced the expression of *Nr4a1* mRNA in all myeloid populations (Fig. 2b). Mice with myeloid-specific deletion of *Nr4a1* achieved via either *LysM*-Cre or *Csf1r*-Cre developed substantially exacerbated EAE compared with that of their *Nr4a1*^{fl/fl} (control) littermates (Fig. 2c,d and Supplementary Fig. 2e). To assess the possibility of a role of Nr4a1 in T cells in EAE onset, we transferred *Nr4a1*^{-/-} 2D2 autoimmune T cells into wild-type mice. In contrast to the myeloid deletion of *Nr4a1*, T cell-specific loss of *Nr4a1* had no effect on disease outcome (Supplementary Fig. 2f). Collectively, these data suggested that Nr4a1 expression in infiltrating monocytes and monocyte-derived macrophages might serve a protective role in EAE.

To further characterize the phenotype of *Nr4a1*^{-/-} mice, we sought to determine the identity of inflammatory cells in their CNS by flow cytometry. To track myeloid cells *in vivo*, we used *Nr4a1*^{+/+} and *Nr4a1*^{-/-} mice expressing a CX3CR1-GFP reporter, in which a GFP-encoding sequence is knocked into one allele of the gene encoding the chemokine receptor CX3CR1, a marker of macrophages and microglia^{23–26}. In this setting, microglia were characteristically CD45^{lo}CX3CR1-GFP^{hi}, while infiltrating monocytes-macrophages were CD45^{hi}CX3CR1-GFP^{lo}CD11b^{hi} (Fig. 2e). To track CNS-infiltrating 2D2 T cells, we used 2D2 cells expressing the red fluorescent protein DsRed²⁷ for injection into recipient mice. The majority (68%) of the inflammatory CNS infiltrate in wild-type mice was 2D2 cells, while 24% of the infiltrate was infiltrating CD11b^{hi} myeloid cells (Fig. 2e). The infiltration of total leukocytes (all CD45⁺) and specifically 2D2 T cells was about sixfold higher in *Nr4a1*^{-/-} mice than in *Nr4a1*^{+/+} mice (Fig. 2e). Among the infiltrating myeloid cells, macrophages represented the main cell population (~80%), and their numbers were significantly greater in *Nr4a1*^{-/-} mice than in *Nr4a1*^{+/+} mice (Fig. 2e).

The number of infiltrating monocytes was also greater in *Nr4a1*^{-/-} mice than in *Nr4a1*^{+/+} mice (Supplementary Fig. 3a). Most of the infiltrating monocytes in both strains were inflammatory Ly6C⁺ monocytes (Supplementary Fig. 3a). Non-classical Ly6C⁻ monocytes

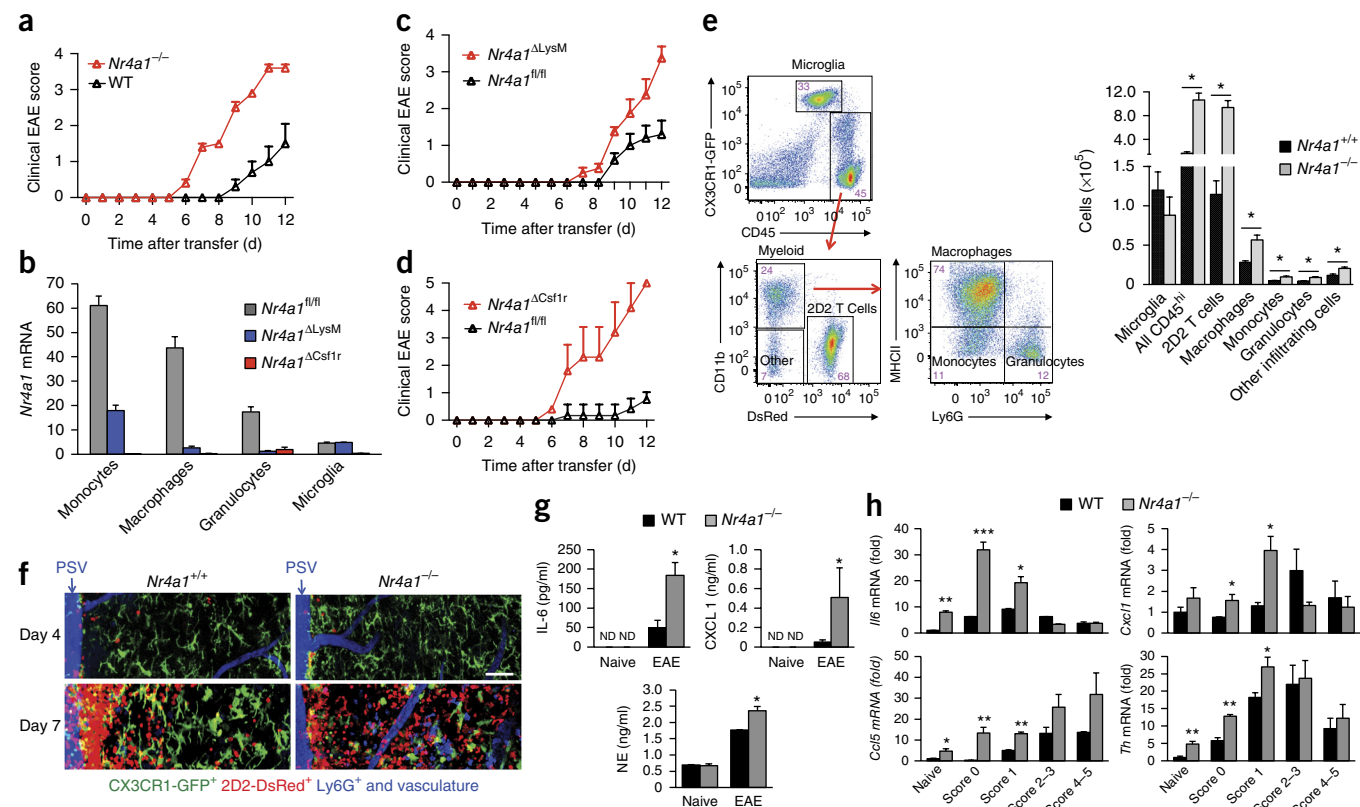


Figure 2 Loss of *Nr4a1* leads to accelerated and exacerbated EAE. (a) Clinical EAE scores of wild-type or *Nr4a1*^{-/-} host mice (*n* = 5 per group) after transfer of 1 × 10⁶ T_H1-polarized 2D2 T cells. *P* < 0.0001 (two-way analysis of variance (ANOVA)). (b) Quantitative RT-PCR analysis of *Nr4a1* mRNA in monocytes, macrophages and granulocytes (sorted from splenocytes) and microglia (sorted from brain cells) of *Nr4a1*^{fl/fl}, *Nr4a1*^{ΔLysM} or *Nr4a1*^{ΔCsf1r} mice (*n* = 4 per group) at baseline, before EAE induction. (c) Clinical EAE scores of *Nr4a1*^{fl/fl} and *Nr4a1*^{ΔLysM} host mice (*n* = 5 per group) after transfer of 1 × 10⁶ T_H1-polarized 2D2 T cells. *P* < 0.05 (two-way ANOVA). (d) Clinical EAE scores of *Nr4a1*^{fl/fl} or *Nr4a1*^{ΔCsf1r} host mice (*n* = 5 per group) after transfer of 0.5 × 10⁶ T_H1-polarized 2D2 T cells. *P* < 0.005 (two-way ANOVA). (e) Isolation and identification of leukocytes from CNS of *Nr4a1*^{+/+} or *Nr4a1*^{-/-} host mice on the CX3CR1-GFP background (*n* = 3 mice per group), on day 7 after transfer of 2 × 10⁶ DsRed⁺ 2D2 T cells: left, gating scheme; right, abundance of cells in various subsets (horizontal axis), assessed by flow cytometry. Numbers in outlined areas (left) indicate percent CX3CR1-GFP^{hi}CD45^{lo} microglia (top) or CX3CR1-GFP^{lo}CD45^{hi} cells (bottom) (top left plot); CD11b^{hi}DsRed⁻ myeloid cells (top left), CD11b⁻DsRed⁻ cells (bottom left) or CD11b⁻DsRed^{hi} 2D2 T cells (bottom right) (bottom left plot); or MHCII⁺Ly6G⁻ macrophages (top left), MHCII⁺Ly6G⁺ granulocytes (bottom right) or MHCII⁺Ly6G⁻ monocytes (bottom left) (bottom right plot). (f) *In vivo* confocal imaging of neutrophils and blood vessels (blue) in the spinal cord adjacent to the posterior spinal vein (PSV) of *Nr4a1*^{+/+} or *Nr4a1*^{-/-} host mice on the CX3CR1-GFP background after transfer of 2 × 10⁶ DsRed⁺ 2D2 T cells, given injection of allophycocyanin-conjugated antibody to Ly6G just before analysis. Scale bar, 50 μm. (g) Blood concentration of IL-6, CXCL1 and NE during onset of disease (day 7) in host mice after transfer of 1 × 10⁶ 2D2 T cells (EAE; *n* = 3 mice per group) or in naive mice (*n* = 4 per group). ND, not detectable. (h) Expression of *Il6*, *Cxcl1*, *Ccl5* and *Th* mRNA among total mRNA extracted from the lumbar spinal cord of wild-type or *Nr4a1*^{-/-} mice before the induction of EAE disease (Naive) or at various stages of EAE disease progression (horizontal axes); results are presented relative to the values in wild-type naive mice. **P* < 0.05, ***P* < 0.01 and ****P* < 0.001 (unpaired Student's *t*-test). Data are representative of six independent experiments (a), two experiments (b,e), three independent experiments (f) or one experiment (h) or are pooled from two independent experiments (c,d,g; error bars (a–e,g,h), s.e.m.).

represented only a minor fraction in *Nr4a1*^{+/+} mice and were almost completely absent in *Nr4a1*^{-/-} mice (Fig. 2e), in line with the reported role of *Nr4a1* in the differentiation of Ly6C⁻ monocytes¹⁵. Studies have shown that in EAE, Ly6C⁺ inflammatory monocytes are recruited to the CNS and give rise to macrophages, and that infiltrating Ly6C⁺ monocyte and monocyte-derived macrophage populations are critical for EAE development^{23,28–30}. Analysis of monocytes isolated from the CNS of *Nr4a1*-GFP mice showed that *Nr4a1* had its highest expression in Ly6C⁻ monocytes but was also expressed in Ly6C⁺ monocytes during disease (Supplementary Fig. 3b), which suggested an important role for *Nr4a1* in regulating these cell types in EAE.

By intravital microscopy, we confirmed earlier infiltration of leukocytes and more-severe CNS damage in *Nr4a1*^{-/-} mice than in *Nr4a1*^{+/+} mice. At 4 d after T cell transfer, few pioneering DsRed⁺ 2D2 T cells were infiltrating the nervous tissue of the thoracic spinal cord segment in both *Nr4a1*^{+/+} mice and *Nr4a1*^{-/-} mice (Fig. 2f and Supplementary

Fig. 4a). In contrast, at day 7, we found substantial infiltration of DsRed⁺ 2D2 T cells into *Nr4a1*^{-/-} spinal cords, whereas in *Nr4a1*^{+/+} mice, the infiltration of 2D2 cells remained confined to an area near the posterior spinal vein (Fig. 2f, Supplementary Fig. 4b and Supplementary Movies 1 and 2). Vascular permeability in brain was similar in wild-type mice and *Nr4a1*^{-/-} mice (Supplementary Fig. 4c), which indicated that the increase in the infiltration of T cells in the *Nr4a1*^{-/-} mice was not caused by impaired function of the blood-brain barrier.

Microglial activation is associated with morphological changes of the cell body and loss of dendritic 'ramified' projections seen in the homeostatic state³¹. Consistent with that, at day 7 after T cell transfer, ramified microglia were much less abundant and rounded microglia were more abundant in *Nr4a1*^{-/-} mice than in *Nr4a1*^{+/+} mice in the spinal cord (Fig. 2f and Supplementary Fig. 5a) and brain (Supplementary Fig. 6). Flow cytometry analyzing CD44 and major histocompatibility complex (MHC) class II, markers of activated microglia, further confirmed

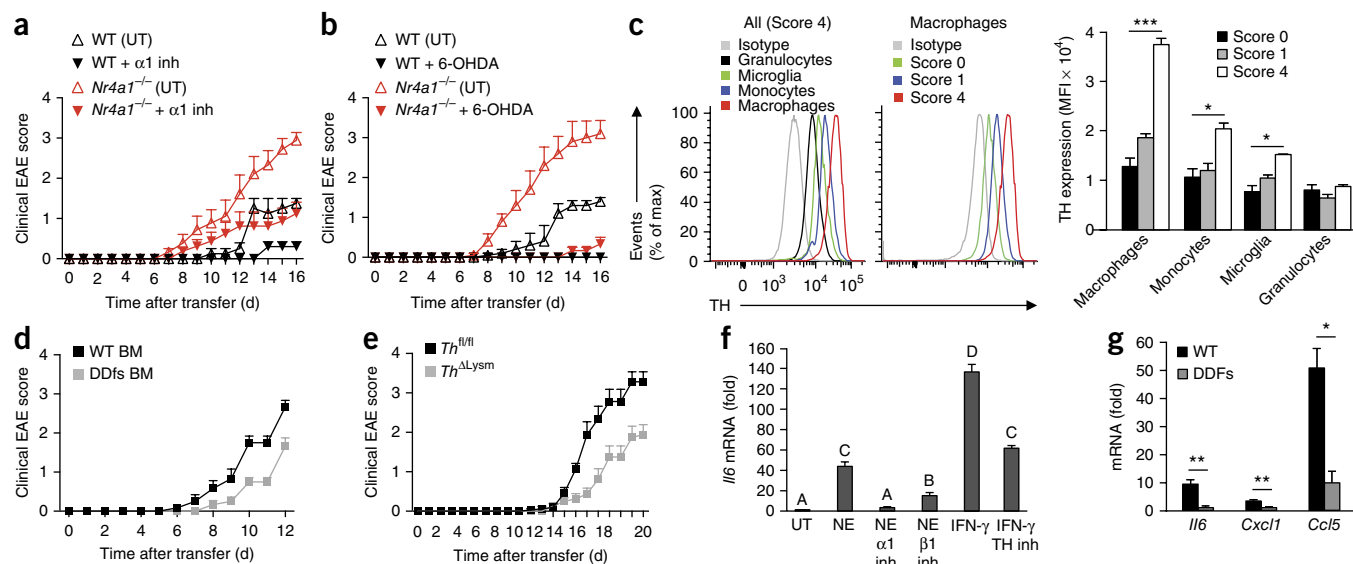


Figure 3 Essential role for NE-producing macrophages in EAE. **(a,b)** Disease progression in wild-type or $Nr4a1^{-/-}$ host mice ($n = 5$ per group) given transfer of 1×10^6 2D2 cells and left untreated (UT) or treated daily with the $\alpha 1$ blocker prazosin ($\alpha 1$ inh) (20 mg per kg body weight) **(a)** or the noradrenergic toxin 6-OHDA (100 mg per kg body weight) **(b)**. $P < 0.01$, wild-type **(a)**; $P < 0.05$, $Nr4a1^{-/-}$ **(a)**; and $P < 0.01$, wild-type and $Nr4a1^{-/-}$ (two-way ANOVA). **(c)** Isolation and identification of CNS-infiltrating TH⁺ cells after transfer of 0.5×10^6 2D2 cells into wild-type host mice (gating as in **Fig. 1b**): left half, flow cytometry analyzing TH in all leukocytes (key) from mice with an EAE score of 4 (left) or macrophages from mice with an EAE score of 0, 1 or 4 (key) (middle); far right, quantification of results at left. **(d)** Disease progression in wild-type host mice ($n = 6$ per group) given transplantation of wild-type or DDfs bone marrow (key) and transfer of 1×10^6 2D2 cells. $P < 0.01$ (two-way-ANOVA). **(e)** Disease progression in $Th^{fl/fl}$ or Th^{ALysM} host mice ($n = 8$ per group) after transfer of 0.5×10^6 2D2 cells. $P < 0.005$ (two-way-ANOVA). **(f)** *Il6* mRNA expression in BMMs left untreated (UT) or treated with 1 μ M NE in the presence or absence of NE receptor blockers ($\alpha 1$ -blocker prazosin or $\beta 1$ -blocker atenolol; 1 μ M each), and in IFN- γ -treated BMM in the presence or absence of the TH inhibitor AMPT (100 μ M); results are presented relative to those of untreated BMMs. Letters above bars indicate differences with a P value of at least <0.001 (one-way ANOVA of multiple comparisons). **(g)** Expression of *Il6*, *Cxcl1* and *Ccl5* mRNA in IFN- γ -treated BMMs from wild-type or DDfs mice (presented as in **f**). $*P < 0.05$, $**P < 0.01$ and $***P < 0.001$ (unpaired Student's t -test). Data are representative of two experiments per treatment **(a,b)** or three independent experiments **(c,d)** or are pooled from two independent experiments **(e)**; error bars **(a–e)**, s.e.m.) or are from one experiment representative of three independent experiments **(f,g)**; error bars, s.e.m. of biological triplicates).

early microglia activation in $Nr4a1^{-/-}$ mice (**Supplementary Fig. 5b,c**). Notably, we also found that at day 7, $Nr4a1^{-/-}$ mice had higher blood concentrations of NE, IL-6 and the IL-6-associated chemokine CXCL1 than did wild-type mice (**Fig. 2g**). Together these data showed that $Nr4a1^{-/-}$ mice developed early and exacerbated EAE, which involved greater induction of NE and IL-6 in blood, accelerated recruitment of autoreactive T cells to the CNS, and early activation of microglia.

Given the higher concentration of NE in $Nr4a1^{-/-}$ mice upon the induction of EAE, we hypothesized that $Nr4a1^{-/-}$ mice had higher expression of TH in their spinal cord than did wild-type mice early in EAE development. To exclude the possibility of changes resulting from the more-advanced disease stage in $Nr4a1^{-/-}$ mice, we compared wild-type mice and $Nr4a1^{-/-}$ mice with the same disease score. As we expected, in the early stages of EAE and even before disease induction, $Nr4a1^{-/-}$ mice showed a greater abundance of mRNA encoding TH, IL-6 and the IL-6-induced chemokines CCL5 and CXCL1 (**Fig. 2h**). This finding was consistent with a published report showing higher TH expression in the CNS of $Nr4a1^{-/-}$ mice³². Our data therefore suggested that $Nr4a1$ regulated the production of NE and IL-6 and thereby limited susceptibility to EAE.

Essential role for NE-producing myeloid cells in EAE

To investigate whether excessive production of NE contributed to the EAE severity of $Nr4a1^{-/-}$ mice, we applied adrenergic inhibitors that block autocrine and paracrine NE signaling. Blockade of $\alpha 1$ adrenergic receptors significantly reduced EAE progression in $Nr4a1^{-/-}$ mice (**Fig. 3a**), but blockade of $\beta 1$ or $\beta 2$ adrenergic receptors did not (**Supplementary Fig. 7a,b**). Depleting mice of catecholamines through use of the noradrenergic neurotoxin 6-OHDA (6-hydroxydopamine)

also inhibited the progression of EAE in both wild-type mice and $Nr4a1^{-/-}$ mice (**Fig. 3b**); this highlighted the critical role of catecholamines in the development of EAE.

To identify the cellular source of NE in EAE, we isolated cells from the CNS of mice with EAE and used flow cytometry to analyze their expression of TH. Notably, we found that TH expression increased with disease progression in macrophages and, to a lesser extent, in monocytes and microglia, but not in granulocytes (**Fig. 3c**). As macrophages represented a principal component of the CNS infiltrate, these results indicated that TH⁺ macrophages recruited to the CNS might be a major driving force behind the development of EAE.

To test that possibility, we used DDfs ('dopamine-deficient floxed stop') mice, which have a non-functional gene encoding TH³³ and in which TH expression was much lower than that of mice with a wild-type gene encoding TH (**Supplementary Fig. 7c**). Wild-type mice given transplantation of DDfs bone marrow were protected from EAE development (**Fig. 3d**). To further substantiate the role of myeloid TH in driving EAE, we used mice with selective deletion of the gene encoding TH in LysM⁺ myeloid cells. These mice had lower TH expression in CNS-infiltrating myeloid cells, but not in microglia or other leukocytes, than that of wild-type mice (**Supplementary Fig. 7d**), and also showed lower concentrations of NE in blood (**Supplementary Fig. 7e**). These mice with myeloid-specific deletion of *Th* were protected from the progression of EAE disease (**Fig. 3e**). The findings that TH expression in granulocytes did not change with disease progression and that *LysM*-Cre-driven deletion of *Th*, as well as transfer of DDfs bone marrow, reduced EAE severity while allowing retention of microglial TH expression, further suggested that TH-expressing

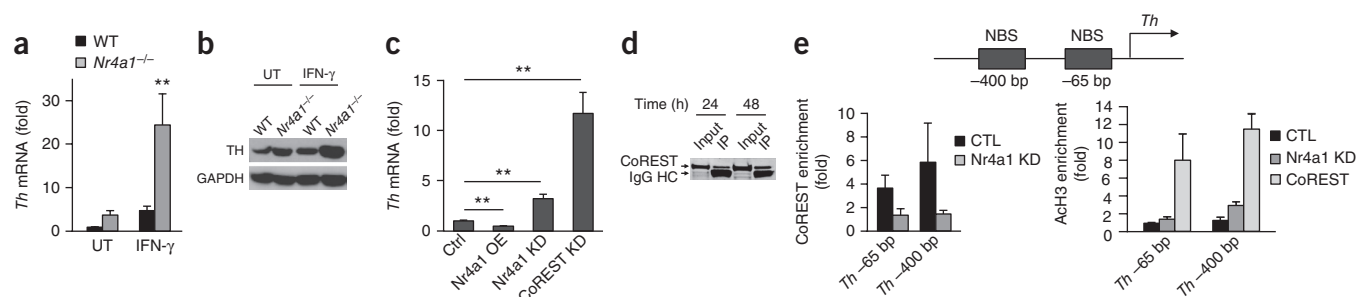


Figure 4 Nr4a1 directly suppresses TH expression in macrophages. **(a)** *Th* mRNA expression in untreated and IFN- γ -treated wild-type or *Nr4a1*^{-/-} BMMs; results are presented relative to those of untreated wild-type BMMs. **(b)** Immunoblot analysis of TH and GAPDH (loading control) in BMMs as in **a**. **(c)** *Th* mRNA expression in control RAW macrophages (Ctrl) (pooled results, far left) and in RAW macrophages with overexpression (OE) of Nr4a1 or after knockdown (KD) of Nr4a1 or CoREST (horizontal axis); results are presented relative to the average of the corresponding control (empty vector (overexpression) or control small interfering RNA (knockdown)). **(d)** Immunoblot analysis of the interaction of Nr4a1 and CoREST in RAW cells at 24 or 48 h after transfection of vector encoding Nr4a1, assessed by immunoprecipitation (IP) with antibody to Nr4a1 and probed with antibody to CoREST. IgG HC, immunoglobulin G heavy chain. Input, immunoblot analysis without immunoprecipitation. **(e)** Chromatin immunoprecipitation analysis of control RAW cells (as in **c**) and RAW cells after knockdown of Nr4a1 or CoREST (key), immunoprecipitated with antibody to CoREST (left) or to acetylated histone H3 (AcH3) (right), followed by quantitative PCR analysis of the immunoprecipitated DNA with the primers for regions 400 base pairs (bp) or 65 base pairs upstream of the transcription start site; results are presented relative to those obtained for a control region (*Pr12b1* promoter). Top, *Th* promoter, with two proximal Nr4a-binding sites (NBS) and primer sites. * $P < 0.05$ and ** $P < 0.01$ (unpaired Student's *t*-test). Data are from one experiment representative of two independent experiments (**a**, **b**, **d**; error bars (**a**), s.e.m. of biological triplicates) or are representative of three independent experiments (**c**; error bars, s.e.m. of biological triplicates) or are pooled from three independent experiments (**e**; error bars, s.e.m.).

infiltrating monocytes and monocyte-derived macrophages were the cells important for EAE development.

Mechanistically, we hypothesized that NE secreted from macrophages during EAE might activate them in an autocrine manner and thereby amplify the neuro-inflammatory cascade, which would lead to the recruitment of T cells to the CNS. To test this hypothesis, we used bone marrow-derived macrophages (BMMs) as a model of monocyte-derived macrophages, since these cells have been shown to both produce NE³⁴ and express adrenergic receptors³⁵. BMMs treated with NE upregulated *Il6*, and this upregulation was abolished by the α 1-adrenergic blocker and, to a lesser extent, by the β 1-adrenergic blocker (Fig. 3f); this suggested that the macrophages were able to sense NE and produce IL-6 in response. *Il6* expression was also reduced by the TH inhibitor AMPT (α -methyl-p-tyrosine) in BMMs treated with interferon- γ (IFN- γ) (Fig. 3f). Upon stimulation with IFN- γ , BMMs from DDf mice showed much lower expression of *Il6* and genes encoding IL-6-driven chemokines than did wild-type BMMs (Fig. 3g), which further confirmed the hypothesis of a role for autocrine catecholamine production in macrophage neuroinflammatory signaling. Collectively, these results suggested that in early stages of EAE, monocytes and macrophages had increased TH activity and secreted NE, which resulted in the production of IL-6 and IL-6-associated chemokines and thus drove the recruitment of leukocytes to the CNS.

Negative regulation of TH by Nr4a1 in macrophages

Given the essential role of macrophage TH in EAE and our observation that the susceptibility of *Nr4a1*^{-/-} mice to EAE was accompanied by higher TH expression, we sought to determine whether Nr4a1 had a relevant role in the regulation of TH in macrophages. Similar to the high Nr4a1 expression *in vivo* in macrophages during neuroinflammation, the expression of both *Nr4a1* mRNA and Nr4a1-GFP was higher in BMMs stimulated with either IFN- γ or NE than in unstimulated BMMs (Supplementary Fig. 8a–c); this suggested a role for Nr4a1 as a sensor of both sympathetic stress and inflammatory stress in macrophages. We therefore sought to determine if Nr4a1 might directly regulate the expression of TH in macrophages as a feedback regulation mechanism. BMMs from *Nr4a1*^{-/-} mice showed much higher expression of *Th* mRNA (Fig. 4a) and TH protein (Fig. 4b) at baseline and after treatment with

IFN- γ than did wild-type BMMs. *Nr4a1*^{-/-} macrophages also secreted significantly more NE than did wild-type BMMs (Supplementary Fig. 8d), and this effect was inhibited by the TH inhibitor AMPT and by 6-OHDA (Supplementary Fig. 8d). Therefore, a lack of Nr4a1 increased TH expression and NE production in macrophages.

To elucidate the molecular mechanism for the regulation of TH by Nr4a1, we used the mouse macrophage cell line RAW264.7 (RAW), as these cells have been shown to both produce NE and sense NE³⁶. Overexpression of *Nr4a1* in RAW cells downregulated the expression of *Th* mRNA and, conversely, knockdown of *Nr4a1* upregulated the expression of *Th* mRNA (Fig. 4c). Moreover, loss of Nr4a1 led to increased activity of a 4.5-kilobase fragment of the mouse *Th* promoter in both BMMs and RAW cells, as assessed by a promoter-reporter assay (Supplementary Fig. 8e). This suggested that Nr4a1 repressed the transcription of *Th*. Collectively, these data identified Nr4a1 as an important negative regulator of TH expression in macrophages.

Another member of the Nr4a family, Nr4a2, has been shown to bind to the *Th* promoter in developing neurons^{37,38} and to downregulate *Th* expression by recruiting the co-repressor CoREST³⁹. CoREST is known to suppress the transcription of neuron-specific genes in a complex with histone deacetylase^{40,41}. Nr4a1 has also been shown to suppress transcription by recruiting the CoREST complex^{42,43}. Because Nr4a1 and Nr4a2 share the same minimal DNA-binding sequence⁷, we hypothesized that Nr4a1 might also downregulate *Th* transcription by recruiting the CoREST complex.

Knockdown of CoREST with small interfering RNA significantly increased *Th* mRNA expression in RAW cells (Fig. 4c). Furthermore, we were able to detect direct interaction between Nr4a1 and CoREST in RAW cells by co-immunoprecipitation (Fig. 4d) and also detected Nr4a1-dependent binding of CoREST to the reported Nr4a-binding sites in the *Th* promoter³⁷ through the use of chromatin immunoprecipitation (Fig. 4e). Chromatin immunoprecipitation analysis also showed increased abundance of acetylated histone H3 at the *Th* promoter after knockdown of *Nr4a1* and, to a greater extent, after knockdown of CoREST (Fig. 4e), which confirmed our hypothesis that CoREST regulates *Th* transcription by recruiting histone deacetylase. Together these results suggested that Nr4a1 directly suppressed the transcription of *Th* in macrophages by recruiting the CoREST complex to the *Th* promoter.

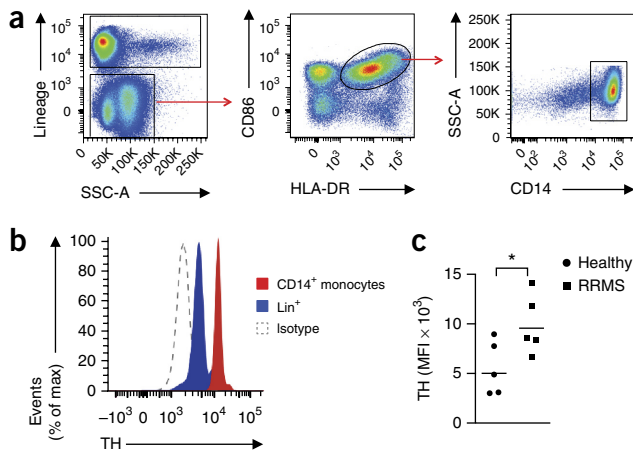


Figure 5 Increased abundance of TH protein in monocytes of patients with MS. (a) Gating strategy for flow cytometry of peripheral blood mononuclear cells from human donors. Monocytes were identified as lineage marker-negative (CD3-CD19-CD66b⁻) (left), HLA-DR⁺ and CD86⁺ (outlined area, middle), and CD14⁺ (outlined area, right). (b) TH expression in monocytes (as in a) and lineage marker-positive cells (Lin⁺) in a patient with relapsing-remitting MS. Isotype, isotype-matched control antibody. (c) TH expression in CD14⁺ monocytes isolated from healthy control donors (Healthy) and patients with relapsing-remitting MS (RRMS), analyzed by flow cytometry. Each symbol represents an individual donor; small horizontal lines indicate the mean. **P* < 0.05 (unpaired Student's *t*-test). Data are from one experiment with one subject (a,b) and or are from two independent experiments with results pooled from five subjects per group (c).

Higher expression of TH in monocytes of patients with MS

Published studies have shown that CNS-infiltrating monocytes and monocyte-derived macrophages are crucial for the pathogenesis of EAE^{23,28–30}. Our data indicated that monocytes and macrophages isolated from CNS during EAE had high expression of TH and that myeloid TH expression was important for the development of EAE. TH expression in human and mouse monocyte-derived macrophages has been reported^{34,44,45}, but its importance in neuroinflammation was not studied. To address the relevance of TH-expressing monocytes and macrophages to human MS, we analyzed peripheral blood mononuclear cells obtained from patients with MS and age- and sex- matched healthy donors (Fig. 5a). We were able to detect TH in both monocytes and lymphocytes in all samples; however, TH expression was higher in monocytes than in all other circulating cells (Fig. 5b). In support of our mouse data, patients with MS exhibited higher TH expression in CD14⁺ monocytes than did healthy control subjects (Fig. 5c). As classical monocytes differentiate into macrophages once they infiltrate the CNS during the onset of EAE^{23,28,30}, these data suggested that TH-expressing monocytes and monocyte-derived macrophages might also have a role in human MS.

DISCUSSION

The main objective of this study was to identify functional and regulatory mechanisms that link stress signaling and neuroinflammation. Adrenergic signaling has long been associated with EAE; however, a role for NE-producing macrophages has not been suggested. A series of publications has shown that prazosin, an α -adrenergic blocker, has a protective effect in both active EAE models and passive EAE models in Lewis rats^{4,5,46,47}, in agreement with our work. Those authors suggested that prazosin might exert its effect by stabilizing the blood-brain barrier; however, since vascular leakage correlated with the EAE score in both control (untreated) rats and treated rats^{46,47}, the decreased leakage through the blood-brain barrier might have been secondary to the overall disease suppression by prazosin. We found no gross

changes in vascular permeability in *Nr4a1*^{-/-} mice; thus, the protective role of Nr4a1 in EAE could not be attributed to such changes. Nevertheless, effects of adrenergic signaling on vascular endothelial cells might indeed serve a role in EAE, such as by inducing IL-6 production and facilitating the recruitment of leukocytes to CNS³. Moreover, our findings indicated that monocytes and macrophages might be a major source of NE for triggering this endothelial activation cascade, in addition to the autocrine signaling we have described.

The main myeloid populations isolated from CNS during EAE included infiltrating monocytes, monocyte-derived macrophages, microglia and granulocytes. Published studies have shown that recruited monocyte-derived macrophages are critical for the development of EAE^{23,28–30}. Macrophages have been shown to express TH and produce catecholamines in mice and humans^{34,36,44,45,48}. In line with those reports, we found a disease-promoting role for TH in recruited macrophages and identified a previously unknown mechanism of neuroprotection by Nr4a1-mediated repression of TH. We observed high expression of Nr4a1 and TH in both infiltrating monocytes and monocyte-derived macrophages and, to a lesser extent, in microglia. There was relatively low expression of both Nr4a1 and TH in granulocytes, which diminished the possibility of Nr4a1-mediated regulation of TH activity in granulocytes. The modulation of EAE in mice with deletion of Nr4a1 and TH driven by Cre expressed via the *LysM*-Cre transgene, which drives deletion in macrophages and monocytes but not in microglia, indicated that microglia did not have a major role in the Nr4a1-mediated neuroprotective mechanism described here. Notably, circulating CD14⁺ monocytes isolated from patients with MS had increased TH expression, which further suggested a disease-promoting role for TH in monocytes and monocyte-derived macrophages. Nevertheless, we cannot exclude the possibility that the lack of patrolling monocytes in *Nr4a1*^{-/-} mice¹⁵ might have contributed to their susceptibility to EAE; although this cell type constitutes a relatively small part of the myeloid infiltrate in the CNS, these cells might serve a protective role by maintaining vasculature homeostasis.

Both proinflammatory roles and anti-inflammatory roles for Nr4a1 have been described in various functional settings, including its modulation of the transcription factor NF- κ B pathway^{49,50}. Another member of the Nr4a family, Nr4a2, expressed in microglia and astrocytes, has been shown to mediate neuroprotection by recruiting the transrepressor CoREST to the promoters of genes that are targets of NF- κ B⁵¹. However, our data demonstrated an alternative neuroprotective function for Nr4a1 in EAE that entailed suppression of NE production in macrophages through recruitment of the CoREST-histone deacetylase complex to the *Th* promoter. Notably, reduced expression of Nr4a1 has been found in peripheral blood mononuclear cells of patients with MS before disease onset^{52,53}, which supports the proposal of a protective role for Nr4a1 in MS. Future studies should further assess the role of Nr4a1-mediated regulation of TH in human monocyte subsets and monocyte-derived macrophages in MS.

Published studies have described the necessity of neurological-immunological communication in controlling inflammation^{3,54}, yet the molecular mechanisms that regulate this are still unclear. Our work has suggested a pivotal role for TH-producing myeloid cells and for neurological-immunological regulators such as Nr4a1 in controlling CNS inflammation. Our work also opens new avenues for potential MS therapies via manipulation of the amount or activity of Nr4a1 to regulate myeloid adrenergic responses.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank D. Metzger (Institut Génétique Biologie Moléculaire Cellulaire) and H. Ichinose (Tokyo Institute of Technology) for *Nr4a1^{fl/fl}* mice; K. Ley for discussions; A. Rao for guidance in composing the manuscript; A. Crotti for guidance on microglia culture; and D. Yoakum for assistance with mouse colony management. Supported by the American Heart Association (13SDG17060117 to I.S. and 12SDG12070005 to R.N.H.), the La Jolla Institute Board of Directors (R.N.H.), Fondation Leducq (M.U.K.), the Sigrid Juselius Foundation (M.U.K.), the Academy of Finland (M.U.K.), the Pacific Northwest Udall Center (P50-NS062684 to M.D.) and the US National Institutes of Health (R01 DK091183-21 to C.K.G. and R01 HL118765 to C.C.H.).

AUTHOR CONTRIBUTIONS

I.S., R.N.H. and H.S. designed, performed and analyzed the experiments; G.C., H.N.N. and R.T. designed and performed the experiments; G.T., R.T., A.B.B., Z.M., S.T., J.M., A.B., M.U.K., S.L.-W.-S. and A.R. performed the experiments; S.S.-A., M.D., G.D.T., A.B.-O., C.K.G., H.B. and C.C.H. designed the experiments; and I.S., R.N.H., H.S. and C.C.H. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Cosentino, M. & Marino, F. Adrenergic and dopaminergic modulation of immunity in multiple sclerosis: teaching old drugs new tricks? *J. Neuroimmune Pharmacol.* **8**, 163–179 (2013).
- Gold, S.M. *et al.* The role of stress-response systems for the pathogenesis and progression of MS. *Trends Immunol.* **26**, 644–652 (2005).
- Arima, Y. *et al.* Regional neural activation defines a gateway for autoreactive T cells to cross the blood-brain barrier. *Cell* **148**, 447–457 (2012).
- Brosnan, C.F. *et al.* Prazosin, an alpha 1-adrenergic receptor antagonist, suppresses experimental autoimmune encephalomyelitis in the Lewis rat. *Proc. Natl. Acad. Sci. USA* **82**, 5915–5919 (1985).
- Brosnan, C.F., Sacks, H.J., Goldschmidt, R.C., Goldmuntz, E.A. & Norton, W.T. Prazosin treatment during the effector stage of disease suppresses experimental autoimmune encephalomyelitis in the Lewis rat. *J. Immunol.* **137**, 3451–3456 (1986).
- Dimitrijević, M. *et al.* Beta-adrenoceptor blockade ameliorates the clinical course of experimental allergic encephalomyelitis and diminishes its aggravation in adrenalectomized rats. *Eur. J. Pharmacol.* **577**, 170–182 (2007).
- Maxwell, M.A. & Muscat, G.E. The NR4A subgroup: immediate early response genes with pleiotropic physiological roles. *Nucl. Recept. Signal.* **4**, e002 (2006).
- Campos-Melo, D. *et al.* Repeated immobilization stress increases *nur77* expression in the bed nucleus of the stria terminalis. *Neurotox. Res.* **20**, 289–300 (2011).
- Chan, R.K., Brown, E.R., Ericsson, A., Kovacs, K.J. & Sawchenko, P.E. A comparison of two immediate-early genes, *c-fos* and *NGFI-B*, as markers for functional activation in stress-related neuroendocrine circuitry. *J. Neurosci.* **13**, 5126–5138 (1993).
- Helbling, J.C., Minni, A.M., Pallet, V. & Moisan, M.P. Stress and glucocorticoid regulation of NR4A genes in mice. *J. Neurosci. Res.* **92**, 825–834 (2014).
- Malkani, S. & Rosen, J.B. Induction of *NGFI-B* mRNA following contextual fear conditioning and its blockade by diazepam. *Brain Res. Mol. Brain Res.* **80**, 153–165 (2000).
- Hamers, A.A. *et al.* Bone marrow-specific deficiency of nuclear receptor *Nur77* enhances atherosclerosis. *Circ. Res.* **110**, 428–438 (2012).
- Hanna, R.N. *et al.* NR4A1 (*Nur77*) deletion polarizes macrophages toward an inflammatory phenotype and increases atherosclerosis. *Circ. Res.* **110**, 416–427 (2012).
- Carlin, L.M. *et al.* Nr4a1-dependent Ly6C^{low} monocytes monitor endothelial cells and orchestrate their disposal. *Cell* **153**, 362–375 (2013).
- Hanna, R.N. *et al.* The transcription factor NR4A1 (*Nur77*) controls bone marrow differentiation and the survival of Ly6C^{low} monocytes. *Nat. Immunol.* **12**, 778–785 (2011).
- Moran, A.E. *et al.* T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. *J. Exp. Med.* **208**, 1279–1289 (2011).
- Kaufman, S. Tyrosine hydroxylase. *Adv. Enzymol. Relat. Areas Mol. Biol.* **70**, 103–220 (1995).
- Bettelli, E. *et al.* Myelin oligodendrocyte glycoprotein-specific T cell receptor transgenic mice develop spontaneous autoimmune optic neuritis. *J. Exp. Med.* **197**, 1073–1081 (2003).
- Sedgwick, J.D. *et al.* Isolation and direct characterization of resident microglial cells from the normal and inflamed central nervous system. *Proc. Natl. Acad. Sci. USA* **88**, 7438–7442 (1991).
- Gautier, E.L. *et al.* Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat. Immunol.* **13**, 1118–1128 (2012).
- Jakubczik, C. *et al.* Minimal differentiation of classical monocytes as they survey steady-state tissues and transport antigen to lymph nodes. *Immunity* **39**, 599–610 (2013).
- Lee, S.L. *et al.* Unimpaired thymic and peripheral T cell death in mice lacking the nuclear receptor *NGFI-B* (*Nur77*). *Science* **269**, 532–535 (1995).
- Ajami, B., Bennett, J.L., Krieger, C., McNagny, K.M. & Rossi, F.M. Infiltrating monocytes trigger EAE progression, but do not contribute to the resident microglia pool. *Nat. Neurosci.* **14**, 1142–1149 (2011).
- Geissmann, F., Jung, S. & Littman, D.R. Blood monocytes consist of two principal subsets with distinct migratory properties. *Immunity* **19**, 71–82 (2003).
- Jung, S. *et al.* Analysis of fractalkine receptor CX3CR1 function by targeted deletion and green fluorescent protein reporter gene insertion. *Mol. Cell. Biol.* **20**, 4106–4114 (2000).
- Saederup, N. *et al.* Selective chemokine receptor usage by central nervous system myeloid cells in CCR2-red fluorescent protein knock-in mice. *PLoS ONE* **5**, e13693 (2010).
- Fan, Z. *et al.* In vivo tracking of 'color-coded' effector, natural and induced regulatory T cells in the allograft response. *Nat. Med.* **16**, 718–722 (2010).
- King, I.L., Dickendesher, T.L. & Segal, B.M. Circulating Ly-6C⁺ myeloid precursors migrate to the CNS and play a pathogenic role during autoimmune demyelinating disease. *Blood* **113**, 3190–3197 (2009).
- Mildner, A. *et al.* CCR2⁺Ly-6C^{hi} monocytes are crucial for the effector phase of autoimmunity in the central nervous system. *Brain* **132**, 2487–2500 (2009).
- Yamasaki, R. *et al.* Differential roles of microglia and monocytes in the inflamed central nervous system. *J. Exp. Med.* **211**, 1533–1549 (2014).
- Aloisi, F. Immune function of microglia. *Glia* **36**, 165–179 (2001).
- Gilbert, F. *et al.* *Nur77* gene knockout alters dopamine neuron biochemical activity and dopamine turnover. *Biol. Psychiatry* **60**, 538–547 (2006).
- Hnasko, T.S. *et al.* Cre recombinase-mediated restoration of nigrostriatal dopamine in dopamine-deficient mice reverses hypophagia and bradykinesia. *Proc. Natl. Acad. Sci. USA* **103**, 8858–8863 (2006).
- Nguyen, K.D. *et al.* Alternatively activated macrophages produce catecholamines to sustain adaptive thermogenesis. *Nature* **480**, 104–108 (2011).
- Gosselin, D. *et al.* Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
- Brown, S.W. *et al.* Catecholamines in a macrophage cell line. *J. Neuroimmunol.* **135**, 47–55 (2003).
- He, X.B. *et al.* Prolonged membrane depolarization enhances midbrain dopamine neuron differentiation via epigenetic histone modifications. *Stem Cells* **29**, 1861–1873 (2011).
- Sakurada, K., Ohshima-Sakurada, M., Palmer, T.D. & Gage, F.H. *Nurr1*, an orphan nuclear receptor, is a transcriptional activator of endogenous tyrosine hydroxylase in neural progenitor cells derived from the adult brain. *Development* **126**, 4017–4026 (1999).
- Yi, S.H. *et al.* *Foxa2* acts as a co-activator potentiating expression of the *Nurr1*-induced DA phenotype via epigenetic regulation. *Development* **141**, 761–772 (2014).
- Andrés, M.E. *et al.* CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc. Natl. Acad. Sci. USA* **96**, 9873–9878 (1999).
- Ballas, N. *et al.* Regulation of neuronal traits by a novel transcriptional complex. *Neuron* **31**, 353–365 (2001).
- Nowyhed, H.N. *et al.* The nuclear receptor *nr4a1* controls CD8 T cell development through transcriptional suppression of *runx3*. *Sci. Rep.* **5**, 9059 (2015).
- Palumbo-Zerr, K. *et al.* Orphan nuclear receptor NR4A1 regulates transforming growth factor-beta signaling and fibrosis. *Nat. Med.* **21**, 150–158 (2015).
- Gaskill, P.J., Carvallo, L., Eugenini, E.A. & Berman, J.W. Characterization and function of the human macrophage dopaminergic system: implications for CNS disease and drug abuse. *J. Neuroinflammation* **9**, 203 (2012).
- Qiu, Y. *et al.* Eosinophils and type 2 cytokine signaling in macrophages orchestrate development of functional beige fat. *Cell* **157**, 1292–1308 (2014).
- Claudio, L. & Brosnan, C.F. Effects of prazosin on the blood-brain barrier during experimental autoimmune encephalomyelitis. *Brain Res.* **594**, 233–243 (1992).
- Goldmuntz, E.A., Brosnan, C.F. & Norton, W.T. Prazosin treatment suppresses increased vascular permeability in both acute and passively transferred experimental autoimmune encephalomyelitis in the Lewis rat. *J. Immunol.* **137**, 3444–3450 (1986).
- Flierl, M.A. *et al.* Phagocyte-derived catecholamines enhance acute inflammatory injury. *Nature* **449**, 721–725 (2007).
- Li, L. *et al.* Impeding the interaction between *Nur77* and *p38* reduces LPS-induced inflammation. *Nat. Chem. Biol.* **11**, 339–346 (2015).
- McMorrow, J.P. & Murphy, E.P. Inflammation: a role for NR4A orphan nuclear receptors? *Biochem. Soc. Trans.* **39**, 688–693 (2011).
- Saijo, K. *et al.* A *Nurr1*/CoREST pathway in microglia and astrocytes protects dopaminergic neurons from inflammation-induced death. *Cell* **137**, 47–59 (2009).
- Achiron, A., Feldman, A. & Gurevich, M. Characterization of multiple sclerosis traits: nuclear receptors (NR) impaired apoptosis pathway and the role of 1-alpha 25-dihydroxyvitamin D3. *J. Neurol. Sci.* **311**, 9–14 (2011).
- Achiron, A. *et al.* Microarray analysis identifies altered regulation of nuclear receptor family members in the pre-disease state of multiple sclerosis. *Neurobiol. Dis.* **38**, 201–209 (2010).
- Riol-Blanco, L. *et al.* Nociceptive sensory neurons drive interleukin-23-mediated psoriasisiform skin inflammation. *Nature* **510**, 157–161 (2014).

ONLINE METHODS

Mice. C57BL/6J wild-type mice (000664; JAX), *Nr4a1*^{-/-} mice²² (006187; JAX) (backcrossed for 15 generations) and 2D2 mice (006912; Jax)¹⁸ on a congenic C57BL/6J background were obtained from The Jackson Laboratory. *Nr4a1*^{fl/fl} mice were provided by D. Metzger and H. Ichinose⁵⁵. The *Nr4a1*-GFP reporter mice have been described¹⁶ and are available from The Jackson Laboratory (016617). DDfs mice (provided by M.D.) have a non-functional gene encoding TH because of insertion of a neomycin-resistance cassette in the first intron of the gene encoding TH³³. They were maintained by provision of L-dopamine to the dam until their weaning, then bone marrow was prepared. *Th*^{ALysM} mice were generated by crossing of *LysM*-Cre⁺ mice with *Th*^{fl/fl} mice⁵⁶. Mice were fed a standard rodent chow diet and were housed in microisolator cages in a pathogen-free facility. Mice were euthanized by CO₂ inhalation. All experiments followed guidelines of the La Jolla Institute for Allergy and Immunology Animal Care and Use Committee, and approval for use of rodents was obtained from the La Jolla Institute for Allergy and Immunology according to criteria outlined in the Guide for the Care and Use of Laboratory Animals from the National Institutes of Health.

EAE induction. EAE was induced as described⁵⁷. CD4⁺T cells isolated from 2D2 mice¹⁸ were differentiated under T_H1 conditions, with plate-bound antibody to (anti-) CD3 (17A2; eBioscience) and anti-CD28 (37.51; eBioscience) in the presence of IL-12 (20 U/ml). On day 5 after initial activation, the cells were harvested and then were restimulated overnight with plate-bound anti-CD3 and anti-CD28 (identified above) in the presence of IL-2 (20 U/ml) and IL-18 (25 ng/ml). Subsequently, the cells were harvested and washed with PBS, and 0.5×10^6 to 2×10^6 cells were transferred intraperitoneally into lightly irradiated (400 rads) 8- to 12-week-old recipient C57BL/6 mice. For most of the experiments, 1×10^6 cells were injected; however, in some experiments, we used 0.5×10^6 cells due to technical reasons. For injection of dsRed-2D2 cells, 2×10^6 cells were used for better visualization of T cells.

Assignment of clinical disease scores. Mice were monitored daily for the development of EAE according to the following criteria: 0, no disease; 1, limp tail; 2, hind limb weakness or partial paralysis; 3, complete hind limb paralysis; 4, front and hind limb paralysis; 5, moribund state. In addition, the body weight of the mice was measured during each clinical assessment⁵⁷. Researchers performing the clinical scoring were blinded to mouse identity.

Blood analysis. The cytokines IL-6 and CXCL1 were measured in the serum with a multiplex inflammatory cytokine/chemokine kit (Meso Scale Discovery). Norepinephrine (NE) was measured with an ELISA kit (Rocky Mountain Diagnostics).

Isolation of CNS-infiltrating cells. Recipient mice were euthanized 15 d after transfer of 2D2 cells. The brain and spinal cord were removed and were digested for 60 min with Liberase TL (Roche). Released cells were passed through a 70- μ m nylon mesh and were subjected to density-gradient centrifugation on a 30–70% Percoll gradient. Mononuclear cells at the interface were harvested and washed extensively, and were re-stimulated for 6 h with PMA (phorbol 12-myristate 13-acetate) (10 nM) and ionomycin (1 μ M) for assessment of cytokine production as described⁵⁷.

Bone marrow macrophage culture. For the generation of bone marrow-derived macrophages (BMM), bone marrow cells were plated in 10-cm bacteriological plastic plates with 10% FCS in RPMI-1640 medium supplemented with 50 ng/ml of recombinant mouse macrophage colony-stimulating factor. On day 7, adherent cells were collected and were replated at a density of 1×10^6 cells per ml in six-well tissue culture plates.

RAW cell culture. RAW264.7 cells were obtained from American Type Culture Collection and absence of mycoplasma was confirmed. Cells were cultured at 37 °C under 5% CO₂ in DMEM supplemented with 10% FBS, penicillin (100 U/ml) and streptomycin (100 mg/ml). For overexpression, cells were transfected with *Nr4a1* cDNA expression vector (Origon) through the use of a Fugene kit (Promega). For knockdown, 0.5 μ g of small interfering RNA specific for *Nr4a1* (Life Technologies) or CoREST (OriGene) was delivered

by a Fugene kit (Promega); as a negative control, 0.5 μ g of OriGene small interfering RNA with a scrambled sequence was used.

Chromatin immunoprecipitation. Chromatin immunoprecipitation was performed as described⁵⁸, with some changes. Cells were fixed for 10 min with 1% formaldehyde, then were lysed and sonicated in sonication buffer (0.1% SDS, 10 mM Tris-HCl, pH 8, and 1 mM EDTA supplemented with protease inhibitors and 5 mM sodium butyrate) for the production of fragments ~500 base pairs in length through the use of a Covaris system. Immunoprecipitation was performed overnight in sonication buffer supplemented with 1% Triton-20 and 150 mM NaCl, with 5 μ g of anti-CoREST (07-455; Millipore) or antibody to acetylated histone H3 (ab4441; Abcam). The DNA recovered was analyzed by quantitative PCR with primers corresponding to two proximal sites previously confirmed to bind *Nr4a2* in the rodent *Th* promoter³⁷: NBS1 (forward, 5'-AGAGGATGCGCAGGAGGTAGGAG-3', and reverse, 5'-GTCCCGAGTTCTGTCTCCAC-3'); and NBS2 (forward, 5'-GGGGACTTGAAGACATCCAA-3', and reverse, 5'-CCCAAGGGTTCAT GTTAGGA-3').

Isolation and analysis of human peripheral blood mononuclear cells. After subjects provided informed consent, blood was obtained from both healthy volunteers and patients at McGill University with recent onset of relapsing-remitting MS who had no exposure to systemic treatments in the previous 3 months. Peripheral blood mononuclear cells were isolated by Ficoll separation from the venous blood of healthy donors and patients with recent onset of relapsing-remitting MS and with no exposure to systemic treatments in the previous 3 months, then the cells were cryo-preserved and analyzed by flow cytometry. Lineage markers were CD3 (UCHT1; BioLegend), CD19 (HIB; BioLegend) and CD66b (G10F5; BioLegend).

Intravital microscopy. *Nr4a1*^{+/+} CX3CR1-GFP^{+/+} or *Nr4a1*^{-/-} CX3CR1-GFP^{+/+} mice given transfer of 2D2-DsRed T cells were anesthetized by intraperitoneal injection of a mixture of ketamine (100–150 mg per kg body weight) and xylazine (15 mg per kg body weight) and were placed on a heating pad. The spinal cord at thoracic vertebrae Th10–Th11 was surgically exposed as described^{59,60} and was immobilized on a custom-made stabilizing ring with a sealed glass coverslip 12 mm wide and 0.17 mm thick⁶¹. For visualization of neutrophils and blood volume, 10 μ g of allophycocyanin (APC)-labeled anti-Ly6G (HK1.4; BioLegend) was injected intravenously through retro-orbital plexus. The imaging was performed on an upright resonant scanning Leica SP5 confocal microscope (Leica Microsystems) through a 25 \times water immersion objective (numerical aperture, 0.95; Leica Microsystems) mounted on a piezoelectric nosepiece z-drive (Piezosystem Jena). GFP, DsRed and APC were imaged simultaneously with 488-nm, 543-nm and 633-nm laser lines and stacks of 20–30 images with a voxel size of 0.714 μ m \times 0.714 μ m \times 2.5 μ m were acquired. Image noise was reduced by line averaging during acquisition and median filtering (radius 1–2) after acquisition. All image processing was performed with Fiji (ImageJ) software and SlideBook software (3i), and the images presented were processed by identical procedures. Spectral spillover of DsRed to the APC channel was removed by subtraction of the DsRed channel from the APC channel. In addition, spillover of GFP to the DsRed channel and background were further reduced by thresholding, followed by contrast stretching.

Line-scan analysis of the infiltration of 2D2 T cells was performed on maximum-intensity projection images of spine parenchyma adjacent to the posterior spinal vein (sampled as stripes of 200 μ m \times 550 μ m). The fluorescence intensity of DsRed in lines perpendicular to the vein was added, then were normalized to that of the wild-type sample at day 7 and is presented as a line profile (second-order smoothing with 50 neighbors; GraphPad Prism software). The 'dendricity index' (perimeter² / area \times 4 π)⁶² of GFP⁺ cells in the spinal parenchyma was determined in maximum-intensity projection images after thresholding of the GFP signal and mask creation. Perimeter and area mask statistics were extracted for the index calculation.

Flow cytometry. CNS-infiltrating cells were collected as described above. All samples were washed in Dulbecco's PBS (Gibco) with 2 mM EDTA and were stored on ice during staining and analysis. Cells (2×10^6 to 4×10^6) were resuspended in 100 μ l flow staining buffer (1% BSA (wt/vol) plus 0.1%

(wt/vol) sodium azide in Dulbecco's PBS). Fcγ receptors were blocked for 15 min and surface antigens on cells were stained for 30 min at 4 °C with the following antibodies: anti-CD3e (145-2C11; BD Pharmingen), anti-Ly6G (1A8; BioLegend), anti-Ly6C (HK1.4; BioLegend), anti-CD11b (M1/70; BD Pharmingen), antibody to MHC class II (M5/114.15.2; eBioscience), anti-CD45 (30-F11; BioLegend), anti-MerTK (BAF591; R&D Systems), used with streptavidin–phycoerythrin (PE)–indotricarbocyanine (Cy7) (405206, BioLegend); and anti-CD64 (X54-517.1; BD biosciences). LIVE/DEAD Fixable Dead Cell Stain (Life Technologies) was used for analysis of viability, and forward- and side-scatter parameters were used for exclusion of doublets from analysis.

For intracellular staining of TH, cells were fixed and made permeable with the Cytofix/Cytoperm Fixation/Permeabilization Solution Kit (BD Biosciences). Cells were stained for 30 min at room temperature with primary anti-TH (LNC1; Millipore) or isotype-matched control antibody (sc-3888; Santa-Cruz). Cells were stained for 30 min at 4 °C with secondary conjugated fluorescent secondary antibody (R-phycoerythrin F(ab')₂ Fragment of Goat Anti-Rabbit IgG (H+L); 31864; Life Technologies).

Calculations of frequency were based on live cells as determined by forward and side scatter and viability analysis. Mean fluorescence intensity was quantified, and expression was calculated relative to that of the wild-type control. Cellular fluorescence was assessed with an LSR II or FACSaria II (all from BD Biosciences) and data were analyzed with FlowJo software (version 9.8; TreeStar).

Blood-brain barrier–permeability analysis. Vascular permeability was evaluated by measurement of the extravasation of Evans blue dye (EBD). EBD (30 mg per kg body weight, in a volume of 200 μl) was injected intravenously in mice anesthetized with ketamine and xylazine and was allowed to circulate for 30 min. The mice were then euthanized, their chest was opened, the inferior vena cava was transected, and the vasculature was flushed with 20 ml saline via the right ventricle for removal of excess intravascular dye. The brain and spinal cord were homogenized and were incubated for 24 h at 37 °C in 100% formamide for extraction of EBD. The concentration of extracted EBD was analyzed by spectrophotometry. Correction for the absorbance (optical density) of contaminating heme pigments was performed as described⁶³ with the equation: $A_{620}(\text{corrected}) = A_{620} - (1.426 \times A_{740} + 0.03)$, where A_{620} is the absorbance at 620 nm and A_{740} is the absorbance at 740 nm. Data were calculated as μg EBD per gram tissue.

Quantitative RT-PCR. RNA was isolated with an RNeasy kit (Qiagen), and cDNA was prepared with an Omniscript RT kit (Qiagen). Quantitative PCR was performed on a Light Cycler (Roche) with RT² primers (Qiagen) and SYBR Green dye (Roche). For analysis of genes encoding members of the Nr4a family, we used a Taqman assay (Life Technologies).

Luciferase promoter-reporter assay. For the generation of a TH luciferase reporter, the *Th* promoter (from position –9 to position –4555 relative to the transcription start site) was cloned into the vector pGL4.10(luc2) (Promega). For the assay, cells were seeded in 12-well plates; after overnight incubation, cells were transfected with 0.5 μg TH luciferase reporter vector and 0.2 μg β-galactosidase reporter vector (pCMVβ; Promega), and with other DNA and small interfering RNA (described above), through the use of a Eugene kit (Promega) for RAW cells or an Amaxa Nucleofactor system (Lonza) for BMM. After 24 h, luciferase activity was measured with the Dual-Luciferase Reporter Assay System (Promega).

Co-immunoprecipitation and immunoblot analysis. RAW cells were transfected to express Nr4a1, and samples were collected at 24 h and 48 h after transfection. Nr4a1 was precipitated with anti-Nr4a1 (M-210; Santa-Cruz) and a Dynabeads co-immunoprecipitation kit (Life Technologies), and the samples were analyzed by immunoblot with anti-CoREST (K72/8; Millipore).

Statistical analysis. Data for all experiments were analyzed with Prism software (GraphPad). Unpaired *t*-tests and two-way analysis of variance were used for comparison of experimental groups. *P* values of less than 0.05 were considered significant. The data seemed to be normally distributed, with similar s.d. and error observed between and within experimental groups.

55. Sekiya, T. *et al.* The nuclear orphan receptor Nr4a2 induces Foxp3 and regulates differentiation of CD4⁺ T cells. *Nat. Commun.* **2**, 269 (2011).
56. Jackson, C.R. *et al.* Retinal dopamine mediates multiple dimensions of light-adapted vision. *J. Neurosci.* **32**, 9359–9368 (2012).
57. Bandukwala, H.S. *et al.* Selective inhibition of CD4⁺ T-cell cytokine production and autoimmunity by BET protein and c-Myc inhibitors. *Proc. Natl. Acad. Sci. USA* **109**, 14532–14537 (2012).
58. Shaked, H. *et al.* Chronic epithelial NF-κB activation accelerates APC loss and intestinal tumor initiation through iNOS up-regulation. *Proc. Natl. Acad. Sci. USA* **109**, 14007–14012 (2012).
59. Davalos, D. *et al.* Stable in vivo imaging of densely populated glia, axons and blood vessels in the mouse spinal cord using two-photon microscopy. *J. Neurosci. Methods* **169**, 1–7 (2008).
60. Farrar, M.J. *et al.* Chronic in vivo imaging in the mouse spinal cord using an implanted chamber. *Nat. Methods* **9**, 297–302 (2012).
61. Thornton, E.E., Krummel, M.F. & Looney, M.R. Live imaging of the lung. *Curr. Protoc. Cytom.* **60**, 12.28.1–12.28.12 (2012).
62. Baxter, C.S., Andringa, A., Chalfin, K. & Miller, M.L. Effect of tumor-promoting agents on density and morphometric parameters of mouse epidermal Langerhans and Thy-1⁺ cells. *Carcinogenesis* **12**, 1017–1021 (1991).
63. Wallace, K.L. & Linden, J. Adenosine A2A receptors induced on iNKT and NK cells reduce pulmonary inflammation and injury in mice with sickle cell disease. *Blood* **116**, 5010–5020 (2010).

Proton-gated Ca^{2+} -permeable TRP channels damage myelin in conditions mimicking ischaemia

Nicola B. Hamilton¹, Karolina Kolodziejczyk^{1†}, Eleni Kougioumtzidou¹ & David Attwell¹

The myelin sheaths wrapped around axons by oligodendrocytes are crucial for brain function. In ischaemia myelin is damaged in a Ca^{2+} -dependent manner, abolishing action potential propagation^{1,2}. This has been attributed to glutamate release activating Ca^{2+} -permeable *N*-methyl-D-aspartate (NMDA) receptors^{2–4}. Surprisingly, we now show that NMDA does not raise the intracellular Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) in mature oligodendrocytes and that, although ischaemia evokes a glutamate-triggered membrane current⁴, this is generated by a rise of extracellular $[\text{K}^+]$ and decrease of membrane K^+ conductance. Nevertheless, ischaemia raises oligodendrocyte $[\text{Ca}^{2+}]_i$, $[\text{Mg}^{2+}]_i$ and $[\text{H}^+]_i$, and buffering intracellular pH reduces the $[\text{Ca}^{2+}]_i$ and $[\text{Mg}^{2+}]_i$ increases, showing that these are evoked by the rise of $[\text{H}^+]_i$. The H^+ -gated $[\text{Ca}^{2+}]_i$ elevation is mediated by channels with characteristics of TRPA1, being inhibited by ruthenium red, isopentenyl pyrophosphate, HC-030031, A967079 or TRPA1 knockout. TRPA1 block reduces myelin damage in ischaemia. These data suggest that TRPA1-containing ion channels could be a therapeutic target in white matter ischaemia.

Ischaemia blocks action potential propagation through myelinated axons¹. Electron microscopy² and imaging of dye-filled oligodendrocytes³ show ischaemia-evoked Ca^{2+} -dependent damage to the capacitance-reducing myelin sheaths, which causes loss of action potential propagation. Glutamate receptor block reduces myelin damage and action potential loss^{2–7}, and glutamate evokes a membrane current in oligodendrocytes mediated by AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid)/kainate and NMDA receptors^{2–4}. Thus, oligodendrocyte damage is thought to be excitotoxic: as for neurons in ischaemia, a rise of glutamate concentration⁸ caused by reversal of glutamate transporters in oligodendrocytes and axons^{9,10} activates receptors that raise² oligodendrocyte $[\text{Ca}^{2+}]_i$, thus damaging the cells.

However, although AMPA/kainate and NMDA receptors regulate oligodendrocyte precursor development^{11,12}, these receptors are down-regulated as the cells mature^{13–15}. How can mature oligodendrocytes be damaged excitotoxically, if they express low levels of glutamate receptors? To investigate how oligodendrocyte $[\text{Ca}^{2+}]_i$ is raised in ischaemia, we characterized ischaemia-evoked membrane current and $[\text{Ca}^{2+}]_i$ changes in cerebellar white matter oligodendrocytes.

Solution mimicking ischaemia (see Methods) evoked an increasing inward current in oligodendrocytes (Fig. 1a, b), often with a faster phase that was obscured when responses in many cells were averaged (Fig. 1c). When applied from before the ischaemia, NBQX and D-AP5 reduced the ischaemia-evoked current by 66% (Fig. 1c, d), while mGluR block had no effect (Extended Data Fig. 1a). Preloading for 30 min with the glutamate transport blocker PDC, to prevent ischaemia-evoked glutamate release by reversal of transporters in the white⁹ and grey¹⁶ matter, also reduced the inward current (by 68%, Fig. 1c, d), while blocking other candidate release mechanisms had no effect (Extended Data Fig. 1a). Thus, glutamate release by reversed uptake helps to trigger the ischaemia-evoked current. Notably, however, current flow through glutamate receptors generates only a small fraction of the sustained inward

current evoked by ischaemia, since applying NBQX and D-AP5 from 200 s after ischaemia had started produced only a non-significant 21% suppression of the ischaemia-evoked inward current (Fig. 1d).

In neurons, an ischaemia-evoked inward current triggered by glutamate release, but maintained by non-glutamatergic mechanisms, generates the ‘extended neuronal depolarization’ (END) that evokes neuronal death¹⁷. However, the ischaemia-evoked current in oligodendrocytes was not prevented by removing external Ca^{2+} , nor by gadolinium, which both block the END¹⁷ (Fig. 1d, e), implying a different mechanism maintains the inward current triggered by glutamate.

Unlike in neurons, where ischaemia evokes a conductance increase mediated by ionotropic glutamate receptors¹⁶, ischaemia decreased the conductance of oligodendrocytes (Fig. 1f, g). The suppressed current reversed below the K^+ reversal potential ($E_K = -104$ mV), at -121 mV with 10 mM (Fig. 1f, Extended Data Fig. 1e) and -118 mV with 0.5 mM HEPES (Fig. 1g). This is expected if ischaemia decreases the membrane K^+ conductance, while $[\text{K}^+]_o$ rises (due to Na^+/K^+ pump inhibition throughout the slice) which increases the inward current at all potentials (see Extended Data Fig. 1b, d; Supplementary Information). K^+ -sensitive electrodes showed that $[\text{K}^+]_o$ in the white matter initially rose slowly during ischaemia, but then increased more abruptly, in parallel with the membrane current (Fig. 1h). The peak rise was 2.35 ± 0.13 mM ($n = 12$) in the white matter and 2.48 ± 0.35 mM ($n = 4$) in the adjacent grey matter (where it reflects the anoxic depolarization of neurons¹⁸). The conductance decrease described above produced 32%, while the $[\text{K}^+]_o$ rise produced 68%, of the inward current in oligodendrocytes at -74 mV (see Supplementary Information). Thus, changes in K^+ fluxes generate the ischaemia-evoked inward current.

Could part of the NMDA-evoked inward current in oligodendrocytes⁴ also reflect a $[\text{K}^+]_o$ rise? Extracellular Cs^+ blocked the NMDA-evoked current while intracellular MK-801 had no effect (Extended Data Fig. 2a–d), suggesting that most of the NMDA-evoked current is generated by $[\text{K}^+]_o$ rising, rather than by oligodendrocyte NMDA receptors. Applying NMDA or raising $[\text{K}^+]_o$, and correlating the resulting inward current with the $[\text{K}^+]_o$ rise occurring (see Supplementary Information; Extended Data Fig. 2e, f), we found that at least 49% of the NMDA-evoked current was attributable to the $[\text{K}^+]_o$ rise that it produced. Since mature oligodendrocytes express few NMDA receptors^{13–15}, this presumably reflects NMDA depolarizing neurons or astrocytes in the slice and releasing K^+ .

These data challenge the idea^{2–5} that, during ischaemia, NMDA receptors in mature oligodendrocytes generate a prolonged calcium influx which damages the cells. We therefore investigated the ion concentration changes evoked in oligodendrocytes by activation of NMDA receptors, using Ca^{2+} -, Na^+ - and K^+ -sensitive dyes loaded into cells from the pipette. When 100 μM NMDA was applied to whole-cell clamped cerebellar granule neurons at -74 mV, as expected it evoked an inward current, and raised $[\text{Ca}^{2+}]_i$ and $[\text{Na}^+]_i$ (Extended Data Fig. 3). In contrast, although NMDA evoked an inward current in oligodendrocytes, it generated no $[\text{Ca}^{2+}]_i$ or $[\text{Na}^+]_i$ elevation; indeed $[\text{Na}^+]_i$

¹Department of Neuroscience, Physiology & Pharmacology, University College London, Gower St., London WC1E 6BT, UK. [†]Present address: Department of Psychiatry, University of British Columbia, 2255 Westbrook Mall, Vancouver, British Columbia V6T 1Z3, Canada.

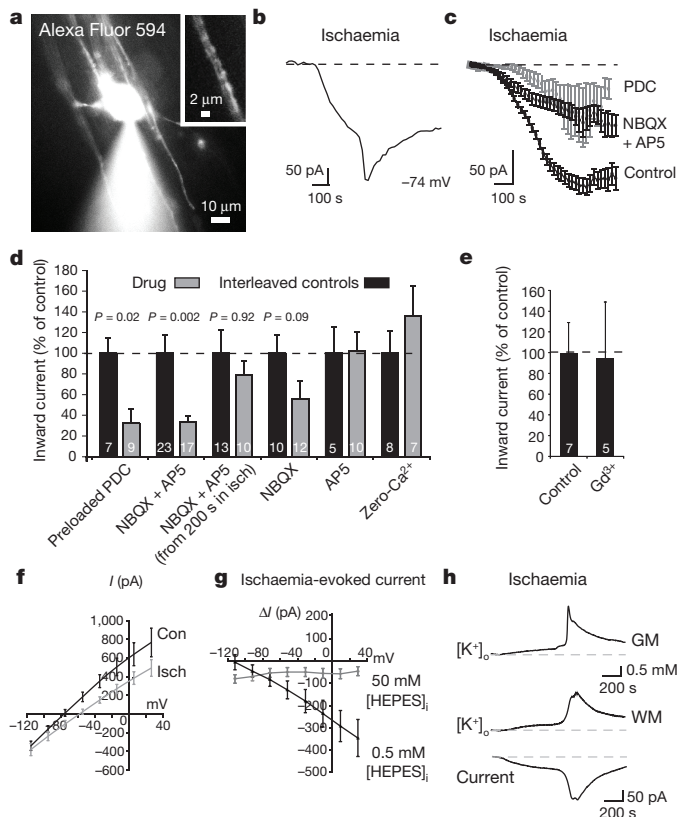


Figure 1 | Ischaemia evokes an inward current in oligodendrocytes by altering K⁺ fluxes. **a**, Whole-cell clamped rat oligodendrocyte. Inset, Alexa dye in processes around an axon. **b**, Ischaemia-evoked membrane current in single cell. **c**, Current in 179 control cells, 12 cells in slices exposed to 25 μ M NBQX and 200 μ M D-AP5 from before ischaemia, or 9 cells in slices preloaded¹⁶ with 1 mM PDC. **d**, Current (normalized to interleaved control cells) from 8–10 min after start of ischaemia in cells preloaded with PDC, exposed to NBQX+AP5 throughout ischaemia or from 200 s after ischaemia starts, or exposed to NBQX or AP5 alone or to zero-Ca²⁺ solution (with 50 μ M EGTA) throughout ischaemia. Mann–Whitney *P* values compare with control cells; cell numbers shown on bars. **e**, Effect of Gd³⁺ (100 μ M) on ischaemia-evoked current at 8–10 min (Mann–Whitney *P* = 0.83). **f**, *I*–*V* relation of 10 cells before and after 5 min ischaemia (10 mM HEPES internal). **g**, Ischaemia-evoked current in 10 cells with 0.5 mM and 9 cells with 50 mM internal HEPES. Ischaemia decreased cell conductance by 2.1 ± 0.7 nS near -70 mV in 11 cells using 10 mM, and by 2.3 ± 0.6 nS in 10 cells using 0.5 mM, internal HEPES; 50 mM HEPES abolished the decrease (Fig. 3i). **h**, Change of [K⁺]_o in grey matter (GM, granule cell layer), and in white matter (WM, different slice) with simultaneously recorded oligodendrocyte current. Error bars, s.e.m.

decreased after applying NMDA (Fig. 2a). NMDA raised [K⁺]_i however. Similar concentration changes were seen at the soma (Fig. 2a) and in the internodal processes where NMDA receptors may be located^{2–4} (Fig. 2b). Like NMDA, raising [K⁺]_o lowered [Na⁺]_i (Fig. 2a, b). A likely explanation is that NMDA raises [K⁺]_o (Fig. 2c), which decreases [Na⁺]_i by activating the Na⁺/K⁺ pump.

The absence of a rise of [Ca²⁺]_i and [Na⁺]_i is surprising if oligodendrocytes express NMDA receptors^{2–4}. Conceivably, NMDA receptors might pass ions into a compartment in their myelinating processes which only certain Ca²⁺-sensing dyes such as X-Rhod-1 can access². However, whether X-Rhod-1 was loaded as an acetoxymethyl ester² or from the pipette, we observed no NMDA-evoked change of [Ca²⁺]_i in the myelinating processes (Fig. 2d). Nevertheless, we could detect spontaneous [Ca²⁺]_i rises propagating through myelinating processes in 55% of oligodendrocytes (Fig. 2e).

To confirm that ischaemia raises² oligodendrocyte [Ca²⁺]_i, we loaded the Ca²⁺-sensing dye Fluo-4 (with Alexa Fluor 594, for ratiometric

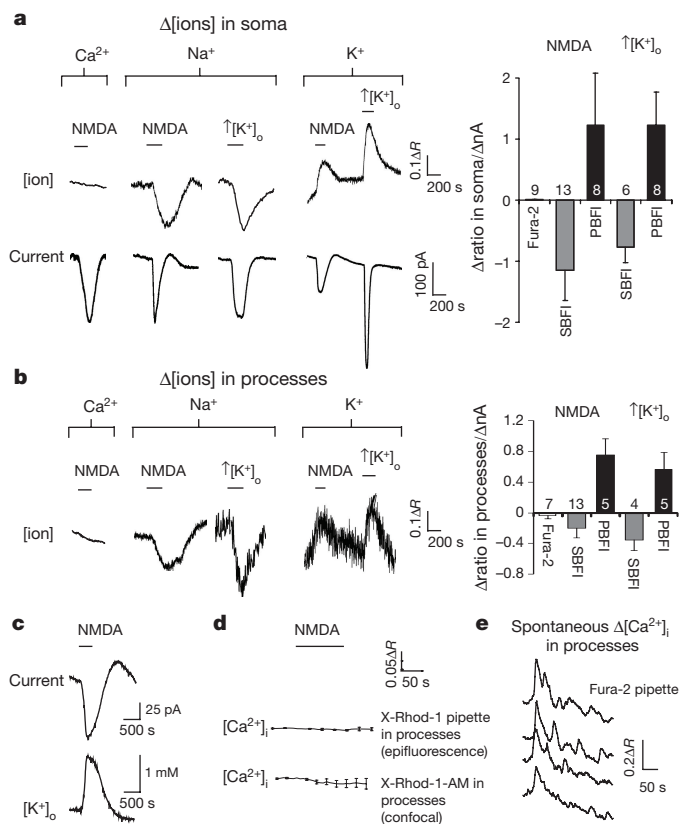


Figure 2 | NMDA does not elevate [Ca²⁺]_i in oligodendrocytes.

a, b, Rat oligodendrocyte membrane current (lower traces, **a**) and background-subtracted fluorescent dye ratio (*R*, see Methods, concentration increases are upwards for all dyes) when measuring [Ca²⁺]_i with Fura-2, [Na⁺]_i with SBFI, and [K⁺]_i with PBFI; 100 μ M NMDA was applied, or [K⁺]_o was raised from 2.5 to 5 mM, with fluorescence measured in soma (**a**) or myelinating processes (**b**). Right panels, peak fluorescence change normalized to evoked current (number of cells on bars) **c**, NMDA-evoked current and simultaneously recorded [K⁺]_o. **d**, Measuring [Ca²⁺]_i with X-Rhod-1, loaded from pipette (*n* = 6) or as an acetoxymethyl ester² (*n* = 15), reveals no NMDA-evoked [Ca²⁺]_i rise. **e**, Spontaneous [Ca²⁺]_i transients in four myelinating processes confirm Fura-2 is working. Error bars, s.e.m.

imaging) into oligodendrocytes from a whole-cell pipette (in current-clamp mode, allowing voltage changes, in case voltage-gated Ca²⁺ channels raise [Ca²⁺]_i). Ischaemia increased [Ca²⁺]_i in the soma and processes over ~10 min. This was abolished if extracellular calcium was removed (Fig. 3a, b), and reduced by removing external K⁺ (Fig. 3c), suggesting that the ischaemia-evoked [K⁺]_o rise promotes calcium entry from the extracellular solution. However, contradicting the earlier report², blocking NMDA receptors with MK-801, D-AP5 and 7-chloro-kynurenat, or blocking NMDA and AMPA/kainate receptors with NBQX and D-AP5 while blocking voltage-gated Na⁺ and Ca²⁺ channels and GABA_A receptors, did not prevent the [Ca²⁺]_i rise (Fig. 3a, b). Similarly, when PDC-preloading reduced transporter-mediated glutamate release, the [Ca²⁺]_i rise was unaffected (Fig. 3c).

Similar experiments using a Mg²⁺-sensitive dye revealed that [Mg²⁺]_i also rises in ischaemia (Fig. 3d). This was not due to ATP breakdown, which releases Mg²⁺, since the [Mg²⁺]_i rise was abolished by removing extracellular Mg²⁺ (Fig. 3d) implying that Mg²⁺ enters across the cell membrane. Surprisingly, ischaemia did not raise [Na⁺]_i (Fig. 3e). Thus ischaemia activates a membrane conductance that allows entry of divalent ions.

Seeking an agent that decreases membrane K⁺ conductance and activates Ca²⁺ entry, we measured the ischaemia-evoked pH change in oligodendrocytes. Ischaemia increased [H⁺]_i on the timescale seen for [Ca²⁺]_i (Fig. 3f). A similar (but smaller) [H⁺]_i rise was evoked by

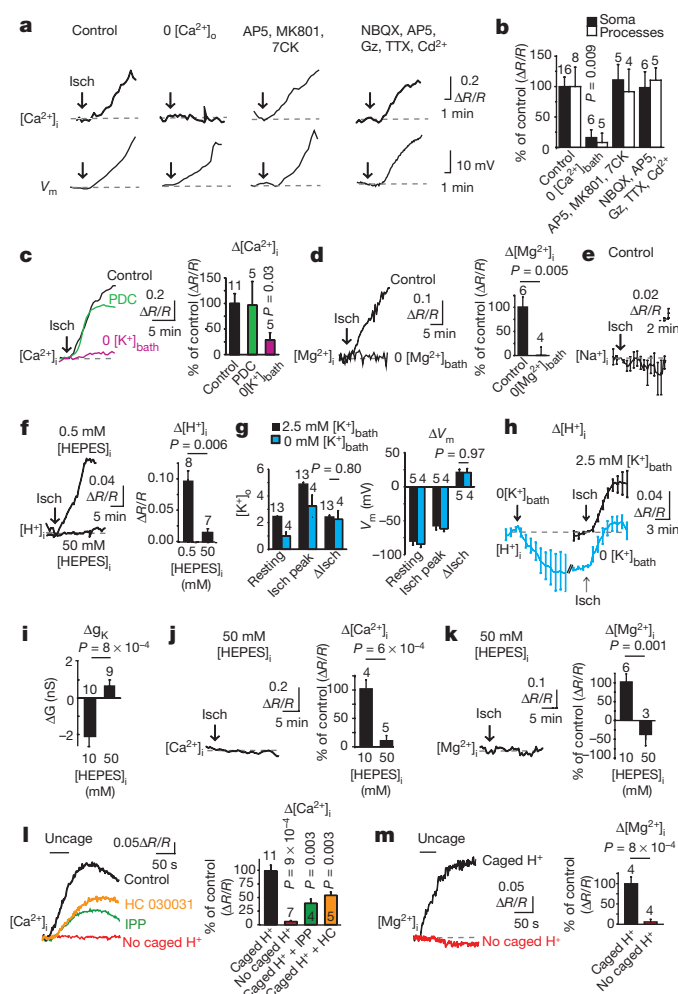


Figure 3 | Ischaemia evokes a $[Ca^{2+}]_i$ and $[Mg^{2+}]_i$ rise gated by internal protons. **a**, Ratiometric Fluo-4/Alexa-Fluor-594 signals ($[Ca^{2+}]_i$) and membrane potential (V_m) in rat oligodendrocytes exposed to ischaemia (starting at arrow), or ischaemia in zero $[Ca^{2+}]_o$, or with drugs at concentrations (μM): AP5 50, MK-801 50, 7-chlorokynurenate (7CK) 100, NBQX 25, GABAzine (Gz) 20, TTX 1, Cd^{2+} (to block Ca^{2+} channels) 100. **b**, Mean data from experiments like **a** (cell numbers on bars; P values compare with soma or process control values). **c**, Data as in **a** after PDC-preloading or in 0 mM $[K^+]_{bath}$. **d**, Ischaemia-evoked $[Mg^{2+}]_i$ rise monitored with Mag-Fluo-4 in normal and Mg^{2+} -free solution. **e**, Ischaemia-evoked $[Na^+]_i$ change monitored with SBFI (6 cells). **f**, Ischaemia-evoked $[H^+]_i$ rise monitored with BCECF with 0.5 mM and 50 mM internal HEPES (P from Mann–Whitney test). **g**, $[K^+]_o$ and oligodendrocyte membrane potential (V_m) with 2.5 or 0 mM $[K^+]_{bath}$, before (Resting) and during ischaemia (Isch peak), and change produced by ischaemia ($\Delta Isch$). **h**, Effect of removing K^+ from bath solution on $[H^+]_i$ in control conditions (relative to value at start of K^+ removal), and $[H^+]_i$ increase evoked by ischaemia (normalized to value at start of ischaemia) with 2.5 or 0 mM bath K^+ . **i**, High [HEPES] $_i$ blocks ischaemia-evoked decrease of membrane conductance. **j**, **k**, Ischaemia-evoked rise of $[Ca^{2+}]_i$ (**j**) and $[Mg^{2+}]_i$ (**k**) are inhibited with 50 mM internal HEPES. **l**, **m**, Uncaging H^+ with light (bars) raises $[Ca^{2+}]_i$ (**l**, an effect reduced by 200 μM IPP or 80 μM HC-030031) and $[Mg^{2+}]_i$ (**m**), but not when caged H^+ is omitted from the pipette. P values in **l** from Mann–Whitney tests. Error bars, s.e.m.

elevated $[K^+]_o$ or NMDA (Extended Data Fig. 4a), suggesting that the ischaemia-evoked $[K^+]_o$ rise partly generates this pH change. To investigate this, we removed external K^+ , which reduced the $[K^+]_o$ in the slice from 2.46 ± 0.02 mM ($n = 13$) to 0.99 ± 0.30 mM ($n = 4$) (Mann–Whitney $P = 0.001$, Fig. 3g), and hyperpolarized the resting potential by 7 mV (-84.0 ± 4.7 to -77.2 ± 4.3 mV, $n = 4$, $P = 0.002$, paired t -test, Fig. 3g). The ischaemia-evoked $[K^+]_o$ rise and depolarization were

unaffected by K^+ removal (Fig. 3g), but the $[H^+]_i$ initially and during ischaemia were reduced (Fig. 3h). These data and those in Fig. 3c suggest that the ischaemia-evoked $[K^+]_o$ rise helps to acidify the cell, which in turn evokes Ca^{2+} entry.

Using an internal solution containing 50 mM HEPES, the ischaemia-evoked rise of $[H^+]_i$ was, as expected, greatly reduced (Fig. 3f). This prevented the decrease of K^+ conductance (Figs 1g, 3i), consistent with intracellular acidity suppressing activity of tonically-active K^+ channels¹⁹. Strikingly, however, buffering intracellular pH also prevented the ischaemia-evoked rise of $[Ca^{2+}]_i$ and $[Mg^{2+}]_i$ (Fig. 3j, k), implying that the ischaemia-evoked $[H^+]_i$ rise activates entry of these cations into the cell. Consistent with this, uncaging protons in oligodendrocytes raised $[Ca^{2+}]_i$ and $[Mg^{2+}]_i$ (Fig. 3l, m).

Few channels allow entry of Ca^{2+} and Mg^{2+} better than Na^+ , but many TRP channels share this property²⁰ (and TRP channel activation can cause ischaemic damage to neurons²¹ and astrocytes²²). Of these channels, only^{20,23,24} TRPA1 and TRPV3 are known to be activated by intracellular H^+ , so we applied modulators of these channels and examined the effect on oligodendrocyte $[Ca^{2+}]_i$ (TRP agonist and blocker specificity are discussed in Supplementary Information). The TRPA1/TRPV3 blocker isopentenyl pyrophosphate²⁵ (IPP) and the TRPA1 blocker²⁶ HC-030031 slowed and reduced the $[Ca^{2+}]_i$ rise evoked by uncaging H^+ in the cell (Fig. 3l). The TRPA1/TRPV3 agonists^{20,27} menthol, vanillin, carvacrol (Cv) and 2-APB all raised $[Ca^{2+}]_i$ in oligodendrocyte somata and myelinating processes, as did the TRPA1 agonists²⁷ AITC, polygodial and flufenamic acid, while the TRPV3 agonists^{20,27} camphor and farnesyl pyrophosphate (FPP) did not (Fig. 4a, b). Thus, TRPA1 subunit-including channels contribute to these responses, but TRPV3 channels are not needed. The carvacrol-evoked rise of $[Ca^{2+}]_i$ was reduced by HC-030031 which blocks TRPA1 but not TRPV3²⁶, by the TRPA1/TRPV3 blocker isopentenyl pyrophosphate²⁵ (IPP), and by TRPA1 knockout (Fig. 4b), again implying involvement of TRPA1 channels. It was unaffected by buffering $[H^+]_i$ (Fig. 4b), consistent with Ca^{2+} entry via TRPA1 being downstream of the $[H^+]_i$ rise, as seen with H^+ uncaging (Fig. 3l).

Using *in situ* hybridization, a TRPA1 probe labelled cerebellar white matter in rat and mouse, while a TRPV3 probe labelled rat only (Extended Data Fig. 5a). Immunocytochemistry revealed that TRPA1- and TRPV3-expressing cells included myelinating oligodendrocytes expressing Olig2 and CC1 (Extended Data Fig. 5b–d).

Consistent with ischaemia raising $[Ca^{2+}]_i$ by activating TRPA1—rather than TRPV3-containing channels, the general TRP blockers^{20,27} ruthenium red (RuR, 10 μM) and La^{3+} (1 mM) and the TRPA1/TRPV3 blocker²⁵ IPP (200 μM) reduced the $[Ca^{2+}]_i$ rise, as did HC-030031 and A967079, which block TRPA1 but not TRPV3^{26,28} (Fig. 4c–e). Knockout of TRPA1 slowed and halved the ischaemia-evoked $[Ca^{2+}]_i$ rise, and the TRPA1/TRPV3 blocker IPP produced no further reduction in the knockout (suggesting no contribution of TRPV3; see Fig. 4f), while knockout of TRPV3 did not affect the $[Ca^{2+}]_i$ rise, and the TRPA1 blocker HC-030031 slowed and halved the rise occurring in the TRPV3 knockout (Fig. 4g). Blockers of many other TRP channels had no effect (Extended Data Fig. 6). Thus TRPA1 is the dominant contributor to the ischaemia-evoked rise of $[Ca^{2+}]_i$ in oligodendrocytes.

The larger (70%) block of the ischaemia-evoked $[Ca^{2+}]_i$ rise by the TRPA1 blocker HC-030031 than by TRPA1 knockout (50%; Figs 4e, f) suggests that there may be compensatory upregulation of another Ca^{2+} entry pathway in the TRPA1 knockout, which normally generates only $\sim 30\%$ of the $[Ca^{2+}]_i$ rise. Introducing high pH-buffering-power solution into the cell blocked the $[Ca^{2+}]_i$ rise in the TRPA1 knockout (Fig. 4h), implying that the non-TRPA1 Ca^{2+} entry pathway is also H^+ -activated. Since the non-specific TRP blockers RuR and La^{3+} abolished the ischaemia-evoked $[Ca^{2+}]_i$ rise, these data suggest that there is another Ca^{2+} -permeable TRP channel (neither TRPA1 nor TRPV3) that is activated by internal H^+ in oligodendrocytes and generates $\sim 30\%$ of the ischaemia-evoked $[Ca^{2+}]_i$ rise.

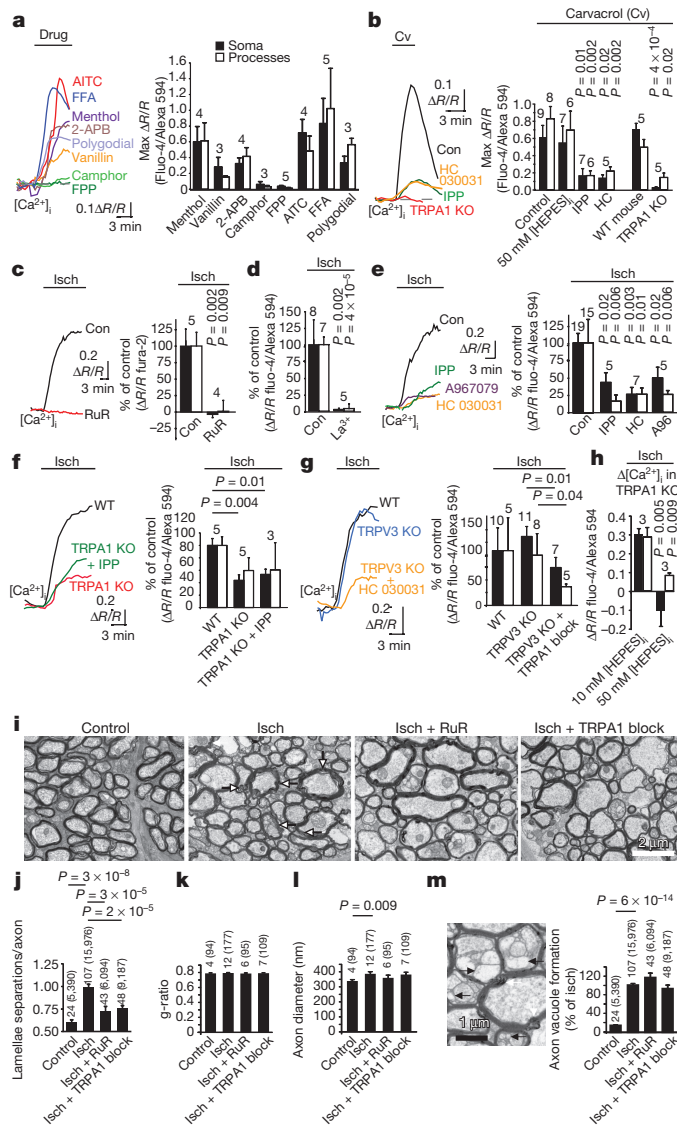


Figure 4 | TRPA1 mediates ischaemic Ca^{2+} accumulation and myelin damage. **a**, $\Delta[Ca^{2+}]_i$ to TRPA1/TRPV3 agonists (mM) (menthol 2, vanillin 1, 2-APB 2), TRPV3 agonists (camphor 2, FPP 0.5) and TRPA1 agonists (AITC 0.5, FFA 1, polygodial 0.2). **b**, $\Delta[Ca^{2+}]_i$ to TRPA1/TRPV3 agonist carvacrol (2 mM, in 1 μ M TTX) is inhibited by TRPA1/TRPV3 antagonist isopentyl pyrophosphate (IPP, 200 μ M, Mann–Whitney test on soma), TRPA1 antagonist HC-030031 (80 μ M), and TRPA1 knockout (KO) (Mann–Whitney test on processes), but not by 50 mM internal [HEPES] (P values compare with control). **c**, **d**, Ischaemia-evoked $\Delta[Ca^{2+}]_i$ is blocked by ruthenium red (RuR, 10 μ M) (**c**), and La^{3+} (**d**, 1 mM, using HEPES-buffered external, Mann–Whitney test on soma). **e**, Block of ischaemia-evoked $\Delta[Ca^{2+}]_i$ by TRPA1/TRPV3 blocker IPP (200 μ M) and TRPA1 blockers HC-030031 (80 μ M) and A967079 (10 μ M). Mann–Whitney P values compare with control. **f**, Ischaemia-evoked $\Delta[Ca^{2+}]_i$ in wild-type mice, with TRPA1 knocked out ($P=0.07$ for processes) and with TRPA1/TRPV3 blocker IPP (200 μ M) also present. **g**, Ischaemia-evoked $\Delta[Ca^{2+}]_i$ in wild-type mice, with TRPV3 knocked out and with TRPA1 blockers HC-030031 (80 μ M, 6 cells) or A967079 (10 μ M, 1 cell) also present (Mann–Whitney P values). **h**, Ischaemia-evoked $\Delta[Ca^{2+}]_i$ in TRPA1 KO with 10 and 50 mM internal [HEPES]. **i**, Electron microscopy showing control optic nerves and myelin decompaction (white arrows) after 60 min ischaemia or ischaemia in RuR (10 μ M), or A967079 (10 μ M) and HC-030031 (80 μ M) together (TRPA1 block). **j–m**, Lamella separations (**j**), g-ratio (**k**), axon diameter (**l**) and axon vacuoles (**m**, electron microscopy shows vacuoles (black arrows) within axon and periaxonal space) in control, ischaemia alone or ischaemia with RuR (10 μ M) or with A967079 (10 μ M) and HC-030031 (80 μ M) (TRPA1 block). Bar numbers are ‘images (axons)’. P values for **j–m** from Mann–Whitney tests, except **l** from Kolmogorov–Smirnov test. Error bars, s.e.m.; data from rat unless stated otherwise.

To assess the role of TRPA1-containing channels in evoking myelin damage, we exposed rat optic nerves to 60 min ischaemia. This led to disruption of myelin sheaths^{2,3}, which we quantified by counting the regions of myelin decompaction (lamellar separation) per axon cross section (see Methods). Taking as baseline the level of decompaction that occurs in control nerves during processing for electron microscopy, ischaemia increased decompaction ($P=3 \times 10^{-8}$), and ruthenium red or the TRPA1 blockers HC-030031 and A967079 (applied together) reduced this increase by 69% ($P=3 \times 10^{-5}$) and 59% ($P=2 \times 10^{-5}$), respectively (Fig. 4i, j). Ischaemia did not affect the axon g ratio (Fig. 4k; see Methods), but increased axon diameter through swelling (Fig. 4l). It also caused some axon vacuolization (Fig. 4m): vacuoles were seen in 4.3% of 2,390 control axons, but in 21% of 15,976 axons after ischaemia. Vacuolization was not prevented by TRP channel block (Fig. 4m), suggesting different mechanisms for axon and myelin damage.

Thus, ischaemic damage to oligodendrocytes differs fundamentally from that in neurons (Extended Data Fig. 7), where $[Ca^{2+}]_i$ is raised by glutamate-gated receptors (and later by TRP channels activated by reactive oxygen species²¹). Contradicting current ideas^{2–4}, ischaemia does not damage oligodendrocytes by activating Ca^{2+} entry through ionotropic glutamate receptors in their membranes. Instead, ischaemia-evoked sodium pump inhibition and glutamate release evoke a long-lasting rise of $[K^+]_o$ that, together with metabolic changes, acidifies the oligodendrocyte, activating H^+ -gated TRP channels through which Ca^{2+} enters.

In the optic nerve, ischaemia-evoked Ca^{2+} entry into oligodendrocytes is blocked by NMDA receptor antagonists², contradicting our demonstrations that NMDA evokes no $[Ca^{2+}]_i$ rise in oligodendrocytes (Fig. 2) and that the ischaemia-evoked $[Ca^{2+}]_i$ rise is unaffected by NMDA receptor blockers (Fig. 3). Conceivably, in the optic nerve, NMDA receptors on astrocytes²⁹ make a greater contribution than in cerebellum to generating the ischaemia-evoked rise of $[K^+]_o$ and thus the $[Ca^{2+}]_i$ rise.

TRPA1 generates ~70% of the ischaemia-evoked $[Ca^{2+}]_i$ rise, and TRPA1 blockers reduce ischaemic damage to myelin (Fig. 4). Consequently, blocking oligodendrocyte TRPA1-containing channels may reduce myelin loss during the energy deprivation that follows stroke, secondary ischaemia caused by spinal cord injury, or hypoxia in multiple sclerosis³⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 September; accepted 7 December 2015.

Published online 13 January 2016.

- Stys, P. K., Ransom, B. R., Waxman, S. G. & Davis, P. K. Role of extracellular calcium in anoxic injury of mammalian central white matter. *Proc. Natl Acad. Sci. USA* **87**, 4212–4216 (1990).
- Micu, I. et al. NMDA receptors mediate calcium accumulation in myelin during chemical ischaemia. *Nature* **439**, 988–992 (2006).
- Salter, M. G. & Fern, R. NMDA receptors are expressed in developing oligodendrocyte processes and mediate injury. *Nature* **438**, 1167–1171 (2005).
- Kárádóttir, R., Cavellier, P., Bergersen, L. H. & Attwell, D. NMDA receptors are expressed in oligodendrocytes and activated in ischaemia. *Nature* **438**, 1162–1166 (2005).
- Bakiri, Y., Hamilton, N. B., Kárádóttir, R. & Attwell, D. Testing NMDA receptor block as a therapeutic strategy for reducing ischaemic damage to CNS white matter. *Glia* **56**, 233–240 (2008).
- McCarran, W. J. & Goldberg, M. P. White matter axon vulnerability to AMPA/kainate receptor-mediated ischemic injury is developmentally regulated. *J. Neurosci.* **27**, 4220–4229 (2007).
- Agrawal, S. K. & Fehlings, M. G. Role of NMDA and non-NMDA ionotropic glutamate receptors in traumatic spinal cord axonal injury. *J. Neurosci.* **17**, 1055–1063 (1997).
- Sarrafzadeh, A. S., Kiening, K. L., Callens, T. A. & Unterberg, A. W. Metabolic changes during impending and manifest cerebral hypoxia in traumatic brain injury. *Br. J. Neurosurg.* **17**, 340–346 (2003).
- Li, S., Mealing, G. A., Morley, P. & Stys, P. K. Novel injury mechanism in anoxia and trauma of spinal cord white matter: glutamate release via reverse Na^+ -dependent glutamate transport. *J. Neurosci.* **19**, RC16 (1999).

10. Back, S. A. *et al.* Hypoxia-ischemia preferentially triggers glutamate depletion from oligodendroglia and axons in perinatal cerebral white matter. *J. Cereb. Blood Flow Metab.* **27**, 334–347 (2007).
11. Yuan, X., Eisen, A. M., McBain, C. J. & Gallo, V. A role for glutamate and its receptors in the regulation of oligodendrocyte development in cerebellar tissue slices. *Development* **125**, 2901–2914 (1998).
12. Lundgaard, I. *et al.* Neuregulin and BDNF induce a switch to NMDA receptor-dependent myelination by oligodendrocytes. *PLoS Biol.* **11**, e1001743 (2013).
13. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
14. De Biase, L. M., Nishiyama, A. & Bergles, D. E. Excitability and synaptic communication within the oligodendrocyte lineage. *J. Neurosci.* **30**, 3600–3611 (2010).
15. Kukley, M., Nishiyama, A. & Dietrich, D. The fate of synaptic input to NG2 glial cells: neurons specifically downregulate transmitter release onto differentiating oligodendroglial cells. *J. Neurosci.* **30**, 8320–8331 (2010).
16. Rossi, D. J., Oshima, T. & Attwell, D. Glutamate release in severe brain ischaemia is mainly by reversed uptake. *Nature* **403**, 316–321 (2000).
17. Limbrick, D. D. Jr, Sombati, S. & DeLorenzo, R. J. Calcium influx constitutes the ionic basis for the maintenance of glutamate-induced extended neuronal depolarization associated with hippocampal neuronal death. *Cell Calcium* **33**, 69–81 (2003).
18. Hamann, M., Rossi, D. J., Mohr, C., Andrade, A. L. & Attwell, D. The electrical response of cerebellar Purkinje neurons to simulated ischaemia. *Brain* **128**, 2408–2420 (2005).
19. Lesage, F. *et al.* TWIK-1, a ubiquitous human weakly inward rectifying K⁺ channel with a novel structure. *EMBO J.* **15**, 1004–1011 (1996).
20. Nilius, B. & Szallasi, A. Transient receptor potential channels as drug targets: from the science of basic research to the art of medicine. *Pharmacol. Rev.* **66**, 676–814 (2014).
21. Aarts, M. *et al.* A key role for TRPM7 channels in anoxic neuronal death. *Cell* **115**, 863–877 (2003).
22. Butenko, O. *et al.* The increased activity of TRPV4 channel in the astrocytes of the adult rat hippocampus after cerebral hypoxia/ischemia. *PLoS One* **7**, e39959 (2012).
23. Wang, Y. Y., Chang, R. B. & Liman, E. R. TRPA1 is a component of the nociceptive response to CO₂. *J. Neurosci.* **30**, 12958–12963 (2010).
24. Cao, X., Yang, F., Zheng, J. & Wang, K. Intracellular proton-mediated activation of TRPV3 channels accounts for the exfoliation effect of α -hydroxyl acids on keratinocytes. *J. Biol. Chem.* **287**, 25905–25916 (2012).
25. Bang, S., Yoo, S., Yang, T. J., Cho, H. & Hwang, S. W. Isopentenyl pyrophosphate is a novel antinociceptive substance that inhibits TRPV3 and TRPA1 ion channels. *Pain* **152**, 1156–1164 (2011).
26. McNamara, C. R. *et al.* TRPA1 mediates formalin-induced pain. *Proc. Natl Acad. Sci. USA* **104**, 13525–13530 (2007).
27. Clapham, D. E. SnapShot: mammalian TRP channels. *Cell* **129**, 220.e1–220.e2 (2007).
28. Chen, J. *et al.* Selective blockade of TRPA1 channel attenuates pathological pain without altering noxious cold sensation or body temperature regulation. *Pain* **152**, 1165–1172 (2011).
29. Hamilton, N. *et al.* Mechanisms of ATP- and glutamate-mediated calcium signaling in white matter astrocytes. *Glia* **56**, 734–749 (2008).
30. Davies, A. L. *et al.* Neurological deficits caused by tissue hypoxia in neuroinflammatory disease. *Ann. Neurol.* **74**, 815–825 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements Supported by a Wellcome Trust Programme Grant and Senior Investigator Award, an ERC Advanced Investigator Award, and the EU (Leukotreat). We thank P. Marcaggi for help with ion-sensitive electrodes, and J. Sexton and J. Wood for knockout mice. K. Kolodziejczyk was in the 4 year PhD Programme in Neuroscience at UCL. We thank L. Arancibia-Carcamo, M. Ford, A. Gibb, R. Jolivet, J. Kittler, M. Sajic, A. Silver and K. Smith for comments on the manuscript.

Author Contributions N.B.H. carried out the experiments in Figs 1g, h, 2, 3 and 4, and Extended Data Figs 2, 3, 4 and 6, and contributed to Figs 1a–d, f and Extended Data Fig. 1. K.K. carried out experiments in Fig. 1a–f and Extended Data Fig. 1. E.K. carried out experiments for Extended Data Fig. 5. All authors designed experiments and analysed data. D.A. and N.B.H. wrote the manuscript. All authors revised the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.B.H. (nicola.hamilton-whitaker@ucl.ac.uk) or D.A. (d.attwell@ucl.ac.uk).

METHODS

Animals. Experiments used Sprague-Dawley rats or transgenic mice of either sex. Data are from rats unless stated otherwise. Animal procedures were carried out in accordance with the guidelines of the UK Animals (Scientific Procedures) Act 1986 and subsequent amendments. TRPV3 knockout (KO) mice were obtained from JAX (<http://jaxmice.jax.org/strain/010773.html>). TRPA1 KO mice were obtained as a double knockout with TRPV1 knocked out (kindly provided by J. Wood and J. Sexton). TRPV1 does not contribute to the ischaemia-evoked $[Ca^{2+}]_i$ rise described here because the TRPV1 antagonist^{20,27} capsazepine did not reduce the ischaemia-evoked $[Ca^{2+}]_i$ rise in rat oligodendrocytes (Extended Data Fig. 6a) and the TRPV1 agonists^{20,27} capsaicin (10 μ M) and camphor (2 mM) did not evoke a $[Ca^{2+}]_i$ rise (see Specificity of drugs acting on TRP channels section and Fig. 4a). Wild-type and (double) KO mice were from a colony obtained by breeding mice doubly heterozygous for the TRPA1 and TRPV1 knockouts. The wild-type and KO mice compared shared the same doubly heterozygous grandparents.

Brain slice preparation. Cerebellar slices (225 μ m thick) were prepared from the cerebellum of P12 rats in ice-cold solution containing (mM) 124 NaCl, 26 NaHCO₃, 1 NaH₂PO₄, 2.5 KCl, 2 MgCl₂, 2–2.5 CaCl₂, 10 glucose, bubbled with 95% O₂/5% CO₂, pH 7.4, as well as 1 mM Na-kynurenate to block glutamate receptors. Slices were then incubated at room temperature (21–24 °C) in the same solution until used in experiments. Cerebellar slices from P10–17 mice were prepared in ice-cold solution containing (mM) 87 NaCl, 25 NaHCO₃, 1.25 NaH₂PO₄, 2.5 KCl, 7 MgCl₂, 0.5 CaCl₂, 25 glucose, 75 sucrose, 1 Na-kynurenate and then transferred to the same solution at 27 °C and allowed to cool naturally to room temperature. Only 1 cell was recorded from in each slice.

Cell identification and electrophysiology. Oligodendrocytes, cerebellar granule cells and hippocampal pyramidal cells were identified by their location and morphology. All cells were whole-cell clamped with pipettes with a series resistance of 8–30 M Ω . Electrode junction potentials were compensated. *I*–*V* relations were from responses to 200 ms voltage steps. Unless otherwise indicated, cells were voltage-clamped at –74 mV.

External solutions. Slices were superfused with either bicarbonate-buffered solution containing (mM) 124 NaCl, 2.5 KCl, 26 NaHCO₃, 1 NaH₂PO₄, 2–2.5 CaCl₂, 1 MgCl₂, 10 glucose, pH 7.4, bubbled with 95% O₂ and 5% CO₂, or with HEPES-buffered solution containing (mM) 144 NaCl, 2.5 KCl, 10 HEPES, 1 NaH₂PO₄, 2–2.5 CaCl₂, 1 MgCl₂, 10 glucose, pH set to 7.3 with NaOH, bubbled with 100% O₂. During experiments when NMDA was applied and ion concentration changes were observed with ion-sensitive dyes, MgCl₂ was omitted from the solution to minimize the Mg²⁺ block. For experiments involving Gd³⁺ and La³⁺, the HEPES-based solution was used and NaH₂PO₄ was omitted. To simulate ischaemia we replaced external O₂ with N₂, and external glucose with 7 mM sucrose, added 2 mM iodoacetate to block glycolysis, and 25 μ M antimycin to block oxidative phosphorylation^{4,31}. All ischaemia experiments were done at 33–36 °C, while applications of NMDA and of TRP channel agonists were at 24 °C. Control and drug conditions were interleaved where appropriate. For calcium imaging experiments when applying ischaemia solution to brain slices from transgenic mice, the experimenter was blind to the genotype.

Intracellular solutions. Cells were whole-cell clamped with electrodes containing either Cs- (to improve voltage uniformity) or K-gluconate-based solution, comprising (mM) 130 Cs-gluconate (or K-gluconate), 2 NaCl, 0.5 CaCl₂, 10 HEPES, 10 BAPTA, 2 NaATP, 0.5 Na₂GTP, 2 MgCl₂, 0.5 K-Lucifer yellow, pH set to 7.2 with CsOH or KOH (all from Sigma). The K⁺-based solution was used for current-clamp experiments. For Ca²⁺ imaging experiments, BAPTA was decreased to 0.01 mM and replaced with 10 mM phosphocreatine, added CaCl₂ was reduced to 10 μ M, and Lucifer yellow was replaced with 1 mM Fura-2, or 200 μ M Fluo-4 with 50 μ M Alexa Fluor 594, or 200 μ M X-Rhod-1 with 50 μ M Alexa Fluor 488 (all from Molecular Probes) to allow ratiometric imaging. For imaging pH, Lucifer yellow was replaced with BCECF (96 μ M) and the HEPES concentration was decreased to 0.5 mM. This [HEPES] was also used for control experiments when examining the effect of 50 mM internal [HEPES] on the ischaemia-evoked current; ischaemia-evoked membrane current changes were indistinguishable when 0.5 and 10 mM HEPES were used (see main text), presumably because endogenous pH buffering dominates at these low [HEPES] levels. For experiments where the pH-buffering capacity of the internal solution was increased, 68 mM K-gluconate and 50 mM HEPES were used. When uncaging protons, 2 mM 1-(2-nitrophenyl) ethyl sulphate sodium salt (NPE-caged protons, Tocris) was added to the pipette solution and 10 mM HEPES was replaced by 30 mM Tris to prevent ultraviolet-light-mediated oxidation³² of HEPES (and K-gluconate was reduced from 130 to 120 mM). For Na⁺ and K⁺ imaging experiments, Lucifer yellow was replaced with 1 mM of the Na⁺-sensing dye SBFI tetra-ammonium salt or of the K⁺-sensing dye PBFI tetra-ammonium salt (Molecular Probes). In some experiments MK-801 (1 mM) was added to the internal solution to block NMDA receptors, and cells

were depolarized to –10 mV for 10 s intermittently over a 20 min waiting period to facilitate MK-801 block of open channels.

Single-cell ion imaging and H⁺ uncaging. For Fura-2, SBFI and PBFI imaging when applying NMDA or during ischaemia, white matter oligodendrocytes and grey matter granule cells were patch-clamped with pipettes containing a solution as described above, fluorescence was excited sequentially at 340 \pm 10 nm and 380 \pm 10 nm, and emitted light was collected at 510 \pm 20 nm. The ratio (*R*) of the emission intensities (340 nm/380 nm), after subtraction of the background intensity averaged over 4 distant areas of the image, was used as a measure of intracellular ion concentration. Increases of ion concentration generated a fall of fluorescence (*F*) excited at 380 nm and a rise in fluorescence excited at 340 nm, which is plotted as $\Delta R/R$ in the graphs shown, with $R = F_{340\text{ nm}}/F_{380\text{ nm}}$; an upward deflection corresponds to a rise of concentration of the sensed ion. For Fura-2, SBFI and PBFI, mean values of *R* before applying NMDA or ischaemia solution were 0.41 \pm 0.05 (*n* = 5), 1.68 \pm 0.14 (*n* = 13) and 1.86 \pm 1.10 (*n* = 8) respectively.

Fluo-4 and Alexa Fluor 594 were used in the internal solution to measure $[Ca^{2+}]_i$ changes ratiometrically during H⁺-uncaging and most ischaemia experiments. To measure $[Mg^{2+}]_i$, Mag-Fluo-4 was used instead of Fluo-4. Fluo-4 (or Mag-Fluo-4) and Alexa Fluor 594 fluorescence were excited sequentially every 2, 10 or 30 s at 488 \pm 10 nm and 585 \pm 10 nm, and emission was collected using a triband filter cube (DAPI/FITC/Texas red, 69002, Chroma). The mean ratio of intensities ($F_{488\text{ nm}}/F_{585\text{ nm}}$) before applying NMDA or ischaemia was 0.81 \pm 0.09 (*n* = 16) for Fluo-4 and 0.55 \pm 0.04 (*n* = 6) for Mag-Fluo-4. Caged-H⁺ were uncaged using 380 \pm 20 nm light for 1 s every 2 s (repeated 30 times) interspersed with the above excitation wavelengths. BCECF was imaged every 30 s at 400 and 480 nm, with emission collected using the above tri-band filter. The ratio (*R*) of the emitted light excited by these two wavelengths ($F_{480\text{ nm}}/F_{400\text{ nm}}$) was used as a measure of $[H^+]_i$ (mean value before ischaemia was 17.6 \pm 0.9, *n* = 8) but, since this ratio decreases with increasing $[H^+]_i$, when plotting changes in $\Delta R/R$ in Fig. 3e and Extended Data Fig. 4 we multiplied them by –1 to produce a trace that increased with $[H^+]_i$.

During ischaemia, slices swelled at the time of the anoxic depolarization. When cells were patch-clamped with calcium dyes, the resulting movement of the cell away from the electrode sometimes caused $[Ca^{2+}]_i$ oscillations within the cells. These oscillations did not occur if the patch pipette was removed (after 2 min to allow dye-filling) before the ischaemic solution was applied. Without the pipette attached to the cell, the time-course of the ischaemia-evoked $[Ca^{2+}]_i$ rise was the same as with the electrode attached, but its amplitude was 69% larger (ratio increase 0.21 \pm 0.03, *n* = 20 versus 0.12 \pm 0.02, *n* = 16). In some experiments (those in Fig. 4c–g and Extended Data Fig. 6) we therefore removed the pipette for calcium-imaging.

Control experiments were carried out to check whether the ischaemia-evoked change of pH would affect our $[Ca^{2+}]_i$ measurements. The internal solution for Ca²⁺-sensing was studied in the experimental bath that the slices usually are placed in. The resting ratio of Fluo-4 fluorescence to Alexa 594 fluorescence was not significantly affected by altering the pH of the solution from 7.05 to 6.55, and this also did not affect the change of ratio produced by adding 200 nM Ca²⁺ to the sensing solution (Extended Data Fig. 4b, c). Thus, even a 0.5 unit pH change occurring in the oligodendrocyte would not significantly affect the calcium dye measurements. **AM dye loading.** X-Rhod-1-AM (38 μ M) dye loading with the myelin marker DIOC6 into P12 cerebellar slices was performed as described previously for optic nerves². Loading times ranged from 1–2 h and a de-esterification period of 30 min at 36 °C was allowed before imaging.

Potassium electrodes. Potassium electrodes were made as described³³. Electrodes were pulled with a resistance of 4–10 M Ω . Electrode tips were silanized by heating them to 250 °C for 7 min while N₂ and *N,N*-dimethyltrimethylsilylamine (Fluka) were gassed into the tip from the back of the electrode. The tip was then filled with either 6% valinomycin, 1.5% potassium tetrakis(4-chlorophenyl)borate (Fluka) and 92.5% 1,2-dimethyl-3-nitrobenzene (Fluka) or the pre-made potassium sensitive ionophore I – cocktail B (Fluka). The electrodes were back-filled with the bicarbonate-buffered external solution mentioned above (2.5 mM K⁺), and attached to a sensitive high resistance electrometer (Model FD 223, World Precision Instruments). A reference electrode tip was placed less than 5 μ m away from the K⁺ electrode tip, and the voltage changes measured by it were subtracted from those measured with the K⁺ electrode. $[K^+]_o$ was determined by calibrating each electrode at the end of every experiment with at least 3 different K⁺ concentrations (1, 2.5, 5, 7.5, 10 or 17.5 mM). To check for cross-reactivity, the $[NaCl]_o$ was decreased by 60 mM which led to a –2.2 \pm 0.1 mV change in voltage (*n* = 3), while a pH change from 7.3 to 6.5 led to a 0.47 \pm 0.22 mV change (*n* = 3). Both of these changes are much less than the 17.5 \pm 0.5 mV change (*n* = 18) seen in response to an increase of $[K^+]_o$ from 2.5 to 5 mM (which is consistent with the electrodes used having an average calibration slope of 60.9 \pm 0.9 mV (*n* = 6) per tenfold change of $[K^+]_o$).

Drugs used. Stock solutions of the following drugs were made up in water: NMDA, AP5, NBQX, MK801, 7-CK, TTX, PDC, IPP, CPG, SKF 96365 and RuR. (S)-MCPG and amiloride were made up in external solution. Carvacrol was made up in ethanol. Bicuculline, bumetanide, HC 030031, A967079, flufenamic acid, capsazepine, FTY720-HCl, 2-APB, AITC, RN1734 and ML204 were made up in DMSO. When used, DMSO and ethanol were also added to control solution at the same concentrations, and did not evoke $[Ca^{2+}]_i$ changes at the concentrations used. Stocks were kept at $-20^{\circ}C$ apart from carvacrol, menthol, vanillin, AITC, and RuR, which were made up fresh on each day of use. To minimise evaporation of carvacrol, vanillin and menthol, lids were kept on until the solutions were used. Gd^{3+} and La^{3+} were applied (as chloride salts) in bicarbonate- and phosphate-free solution to avoid chelation by these anions (see External solutions section earlier).

Immunohistochemical labelling of oligodendrocytes. Cerebellar slices were fixed for 30 min in 4% paraformaldehyde (PFA), and incubated for 1 to 6 h in 0.1% Triton X-100, 10% goat serum in phosphate-buffered saline at $21^{\circ}C$, then with primary antibody at $4^{\circ}C$ overnight with agitation, and then 2 h or overnight at $24^{\circ}C$ with secondary antibody. Primary antibodies were: anti-CC1 (mouse, 1:300, Calbiochem OP80 monoclonal) and anti-Olig-2 (rabbit, 1:700, Millipore, AB9610 polyclonal). Secondary antibodies were: goat anti-rabbit Alexa Fluor 488 or 568 (Molecular Probes, 1:1,000), donkey anti-rabbit Alexa Fluor 488 (Millipore, 1:1,000), and goat anti-mouse Alexa Fluor 568 (Millipore, 1:1,000).

Antibody labelling and *in situ* hybridization for TRPA1 and TRPV3. TRPA1 and TRPV3 antibodies appeared to label the myelinating processes and somata of oligodendrocytes in rat, but the labelling was not significantly different in wild-type mice and mice with TRPA1 or TRPV3 knocked out (data not shown). We therefore turned to *in situ* hybridization.

Solutions used for *in situ* hybridization were pretreated with 0.1% DEPC. Animals were perfused with PBS followed by 4% PFA. Brains were post-fixed in 4% PFA overnight at $4^{\circ}C$, cryoprotected in 20% sucrose overnight at $4^{\circ}C$ and frozen in Tissue-Tek OCT Sections (20 μ m) collected onto Superfrost Plus microscope slides (VWR International) were hybridized at $65^{\circ}C$ overnight with hybridization buffer [50% v/v deionized formamide (Sigma), 10% w/v dextran sulphate (Fluka), 0.1 mg ml $^{-1}$ yeast tRNA (Roche), 1 \times Denhardt's solution (Sigma) and 1 \times 'salts' (200 mM NaCl, 5 mM EDTA, 10 mM Tris-HCl pH 7.5, 5 mM Na $_2$ HPO $_4$, 5 mM Na $_2$ HPO $_4$)] containing digoxigenin (DIG)-labelled antisense RNA probe (1:1,000). Sections were washed with a washing solution (50% v/v formamide, 1 \times SSC, 0.1% Tween 20) three times at $65^{\circ}C$ for 30 min, followed by two 1 \times MABT (100 mM maleic acid, 150 mM NaCl, pH 7.5, 0.1% Tween-20) washes at room temperature for 30 min each. Sections were subsequently blocked with blocking solution (2% w/v blocking reagent (Roche Diagnostics), 10% v/v heat-inactivated sheep serum (Sigma) in 1 \times MABT) for 1 h at room temperature and incubated with anti-DIG antibody conjugated with alkaline phosphatase (AP) (Roche Diagnostics, 1:1,500 in blocking solution) at $4^{\circ}C$ overnight. Sections were then washed in 1 \times MABT 5 times for 20 min each at room temperature, followed by two 5 min washes in staining buffer (100 mM NaCl, 50 mM MgCl $_2$, 100 mM Tris-HCl, pH 9.5, 0.1% Tween-20). Development was performed at $37^{\circ}C$ for 24–48 h overnight with nitroblue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate in freshly prepared staining solution (50% v/v staining buffer, 25 mM MgCl $_2$, 5% w/v polyvinyl alcohol). Sections were washed in PBS and immunohistochemistry was performed as described above. The plasmids used to generate RNA probes were: IMAGE clone 40129486 for *Trpa1* (linearized with ClaI and transcribed with T3 RNA polymerase) and IMAGE clone 40047664 for *Trpv3* (linearized with XhoI and transcribed with SP6 RNA polymerase). *In situ* hybridization was repeated using at least three animals for each probe.

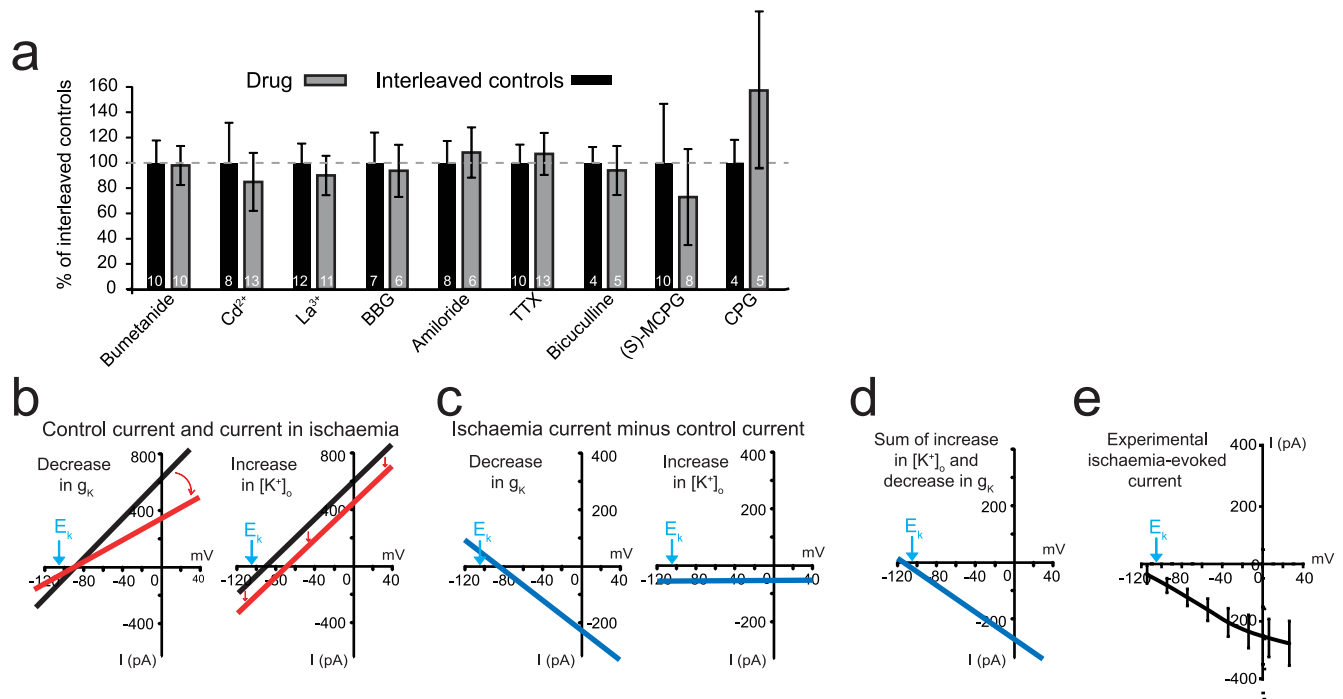
Quantifying myelin decompaction during chemical ischaemia using electron microscopy. For chemical ischaemia experiments, optic nerves were dissected from P28 Sprague-Dawley rats and incubated for 1 h at $36^{\circ}C$ in either control or ischaemic solution with and without the TRPA1/V3 channel blocker ruthenium red (10 μ M) or the combined presence of the TRPA1 blockers HC-030031 (80 μ M) and A967079 (10 μ M). The optic nerves were then immersion fixed in 2% paraformaldehyde and 2% glutaraldehyde in 0.1 M cacodylate buffer overnight.

All samples were then post-fixed in 1% OsO $_4$ /0.1 M cacodylate buffer (pH 7.3) at $3^{\circ}C$ for 2 h before washing in 0.1 M cacodylate buffer (pH 7.3). The samples were dehydrated in a graded ethanol-water series at $3^{\circ}C$ and infiltrated with Agar 100 resin mix. The nerve was then cut transversely at the mid-point, blocked out and hardened. Ultra-thin sections were taken, 300 μ m from the cut end of the middle of the nerve, on a Reichert Ultracut S microtome. Sections were collected on 300 mesh copper grids and stained with lead citrate. The sections were imaged using a Joel 1010 transition electron microscope and a Gatan Orius camera.

In 3 out of 4 experiments the experimenter was blinded to the drug condition before imaging (all 4 experiments gave similar results). One section was used from each nerve and eight 21.5 μ m \times 17.3 μ m images were collected at $\times 8,000$ magnification, four from the peripheral borders of the nerve at 0° , 90° , 180° and 270° positions on the section, and four covering the central portion of the nerve. In all experiments the image identities were then blinded before analysis, and the number of large separations of lamellae (decompaction) was counted. Decompaction was defined as a visible white inter-lamellar gap being present between at least 2 normal lamellae. Regions of decompaction were normally separated from each other by an area of compact myelin, but when most of the myelin surrounding an axon had separated lamellae, decompacted regions were counted at 0.5 μ m intervals around the sheath. The number of decompacted regions was normalized to the number of axons per image. Some decompaction occurred even in control nerves as a result of the processing for electron microscopy, so we assessed drug block of decompaction by quantifying the ischaemia-evoked increase in decompactions seen without and with the drug present. Myelin g ratios were calculated as the square root of the ratio of the area of the axon to the area of the axon plus myelin sheath. When drawing lines around the axon and sheath, areas of decompaction were ignored, that is, we interpolated the lines from regions that were not decompacted. Axon diameter was calculated as $(4(\text{axon area})/\pi)^{0.5}$. Axon vacuolization was defined as the inclusion of one or more large ($>0.1 \mu$ m) empty membrane bound (often circular) organelles within the axon or periaxonal space (Fig. 4m) which may reflect rearrangement of internal axonal membranes or be formed from inclusion of myelin membranes into the axon (Fig. 4m).

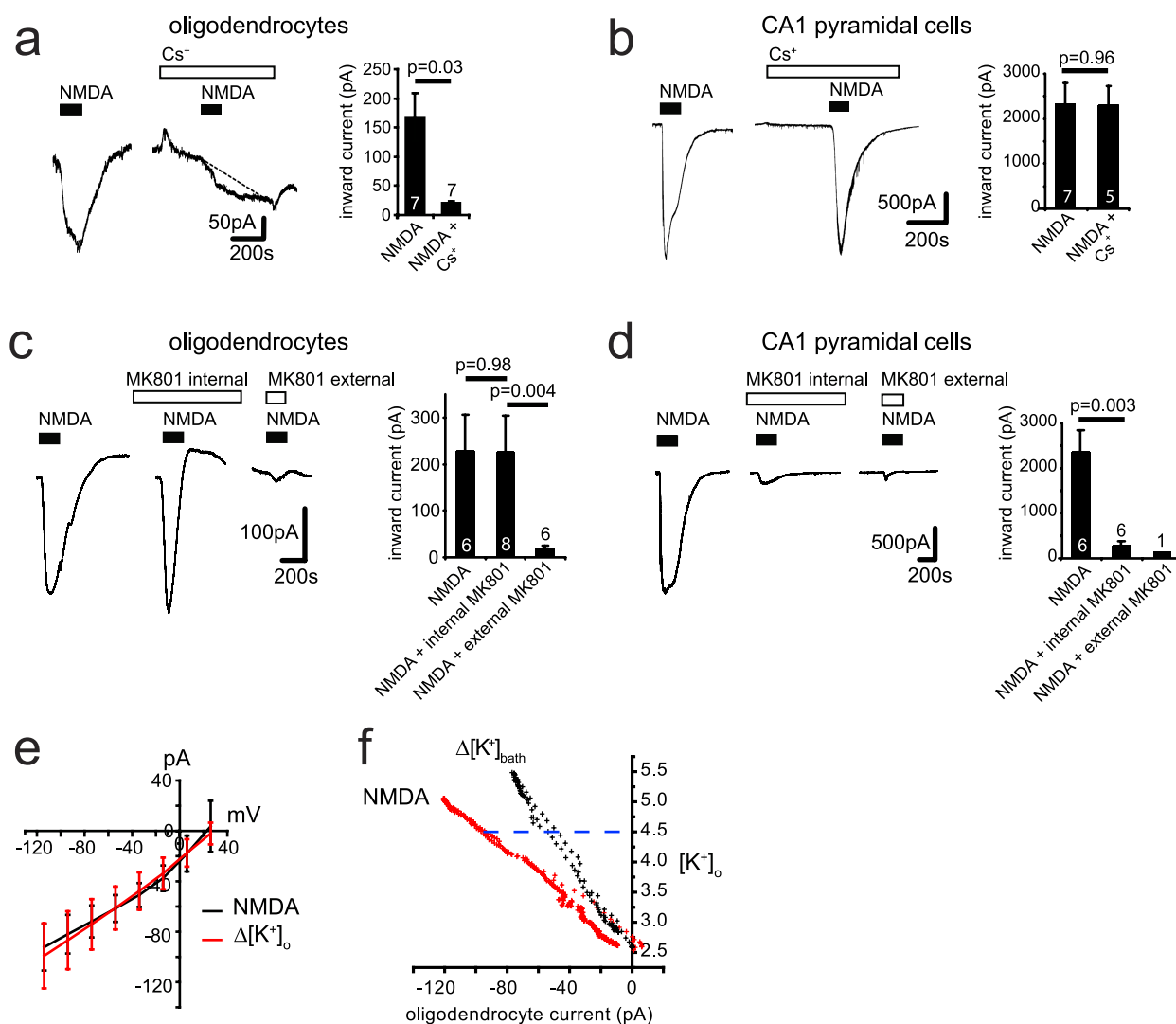
Statistics. Data are presented as mean \pm s.e.m. Experiments were carried out on brain slices from at least 3 animals on at least 3 separate days, except for a few experiments using expensive drugs which were done on only 2 days. Only 1 cell was recorded from in each slice, so the numbers of cells given are also the numbers of slices. *P* values are from two-tailed Student's *t*-tests (for normally distributed data, assessed using Shapiro-Wilk tests) or Mann-Whitney *U* tests (for non-normally distributed data). Normally distributed data were tested for equal variance (*P* < 0.05, unpaired *F*-test) and homo- or heteroscedastic *t*-tests were chosen accordingly. *P* values in the text are from unpaired *t*-tests unless otherwise stated. When small sample sizes (*n* \leq 4) achieved *P* < 0.05, analysis of sample and effect size typically demonstrated a power for detecting the observed effect of 80–99% (mean 92%), with two exceptions: the process data in Fig. 4c (power 78%) and the soma data in Fig. 4h (power 75%). For multiple comparisons within one experiment (usually one figure panel, but measurements of $[Ca^{2+}]_i$ in somata and processes were treated as separate experiments even when plotted in the same figure panel), *P* values were corrected using a procedure equivalent to the Holm-Bonferroni method (for *N* comparisons, the most significant *P* value is multiplied by *N*, the 2nd most significant by *N* – 1, the 3rd most significant by *N* – 2, etc.; corrected *P* values are significant if they are less than 0.05). All statistical analysis was conducted using OriginLab software.

31. Allen, N. J., Kárádóttir, R. & Attwell, D. A preferential role for glycolysis in preventing the anoxic depolarization of rat hippocampal area CA1 pyramidal cells. *J. Neurosci.* **25**, 848–859 (2005).
32. Keynes, R. G., Griffiths, C. & Garthwaite, J. Superoxide-dependent consumption of nitric oxide in biological media may confound *in vitro* experiments. *Biochem. J.* **369**, 399–406 (2003).
33. Marcaggi, P., Jeanne, M. & Coles, J. A. Neuron-glia trafficking of NH $_4^+$ and K $^+$: separate routes of uptake into glial cells of bee retina. *Eur. J. Neurosci.* **19**, 966–976 (2004).



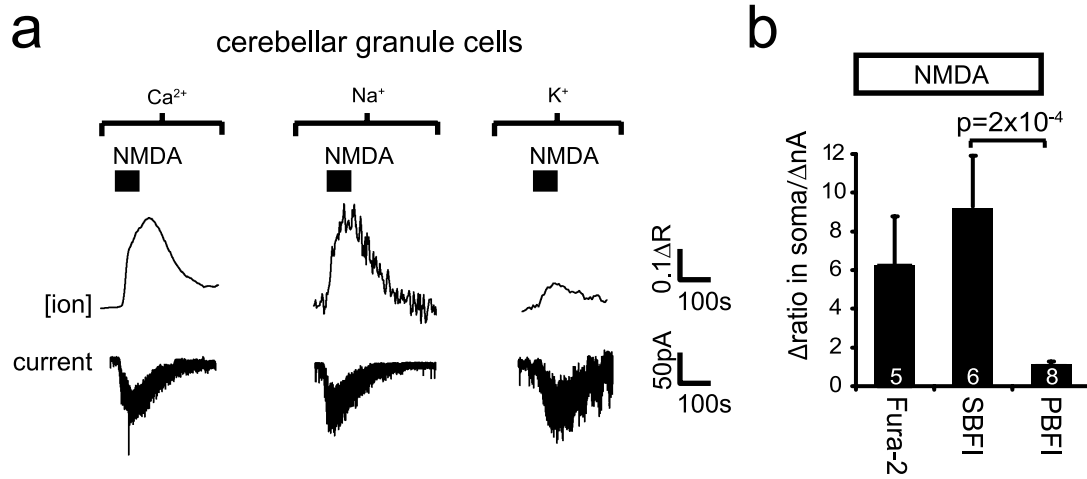
Extended Data Figure 1 | Tests for causes of the ischaemia-evoked current. **a**, Effect of blocking various putative glutamate release mechanisms (blocker concentrations given in Supplementary Information) on peak ischaemia-evoked currents measured in the presence of each drug and in interleaved controls (data from rat). No significant differences were measured ($P > 0.20$). **b**, Schematic showing effect of ischaemia-evoked decrease in resting conductance (which is dominated by g_K , left) and ischaemia-evoked $[K^+]_o$ rise (right) on oligodendrocyte membrane

current. Black lines are control $I-V$ relations. Red lines are $I-V$ relations in ischaemia showing the effect of a conductance decrease (left) or of a positive shift of reversal potential due to $[K^+]_o$ rising (right). **c**, Ischaemia-evoked current change for the two mechanisms in **b** (cf. Fig. 1g). **d**, Sum of currents in **c** gives an $I-V$ relation with a reversal potential more negative than E_K . **e**, Experimentally observed ischaemia-evoked current in 10 oligodendrocytes with 10 mM internal HEPES (difference of curves in Fig. 1f). Error bars are s.e.m.



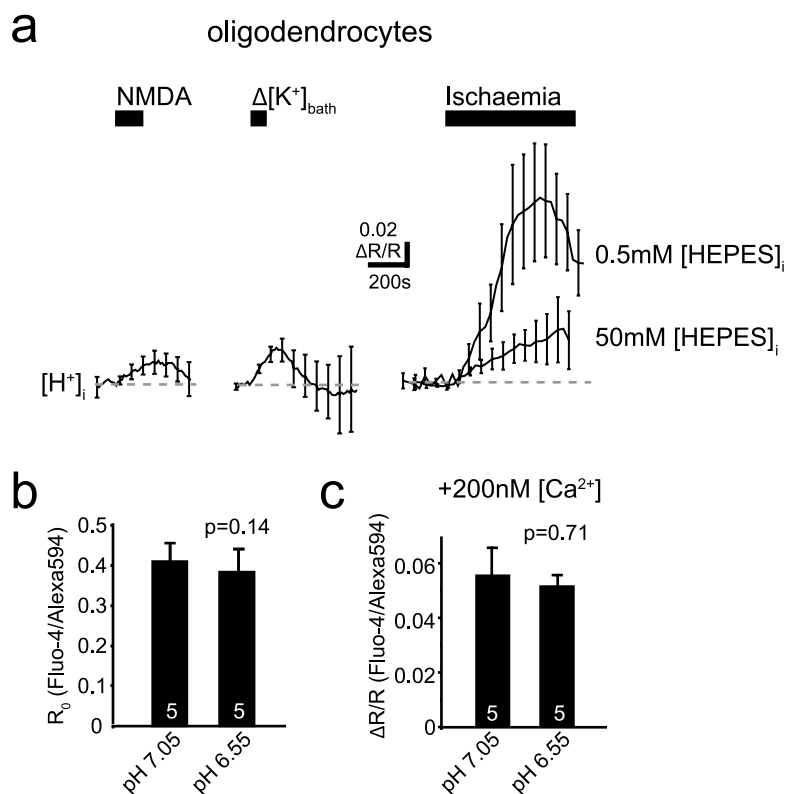
Extended Data Figure 2 | K⁺ flux changes generate the oligodendrocyte NMDA-evoked current. **a, b**, Extracellular Cs⁺ (30 mM, replacing Na⁺) reduces the inward current evoked by 100 μ M NMDA at -74 mV in rat oligodendrocytes (**a**), but not in hippocampal CA1 pyramidal neurons (**b**). **c, d**, Intracellular MK-801 (1 mM) has no effect on NMDA-evoked currents in oligodendrocytes (**c**) but blocks them in pyramidal cells (**d**), while extracellular MK-801 (50 μ M) blocks both. **e**, Voltage-dependence

of the current evoked in 16 oligodendrocytes by 100 μ M NMDA and by elevating $[K^+]_o$ from 2.5 to 5 mM. **f**, Specimen plot of membrane current in an oligodendrocyte versus local $[K^+]_o$ in response to applying 100 μ M NMDA or elevating $[K^+]_o$ from 2.5 to 5 mM. Horizontal cell line shows that if NMDA raises $[K^+]_o$ to (say) 4.5 mM, the current attributable to the $[K^+]_o$ rise alone is 51% of the NMDA-evoked current. Mean value in 11 cells was 49% (see Supplementary Information). Error bars are s.e.m.



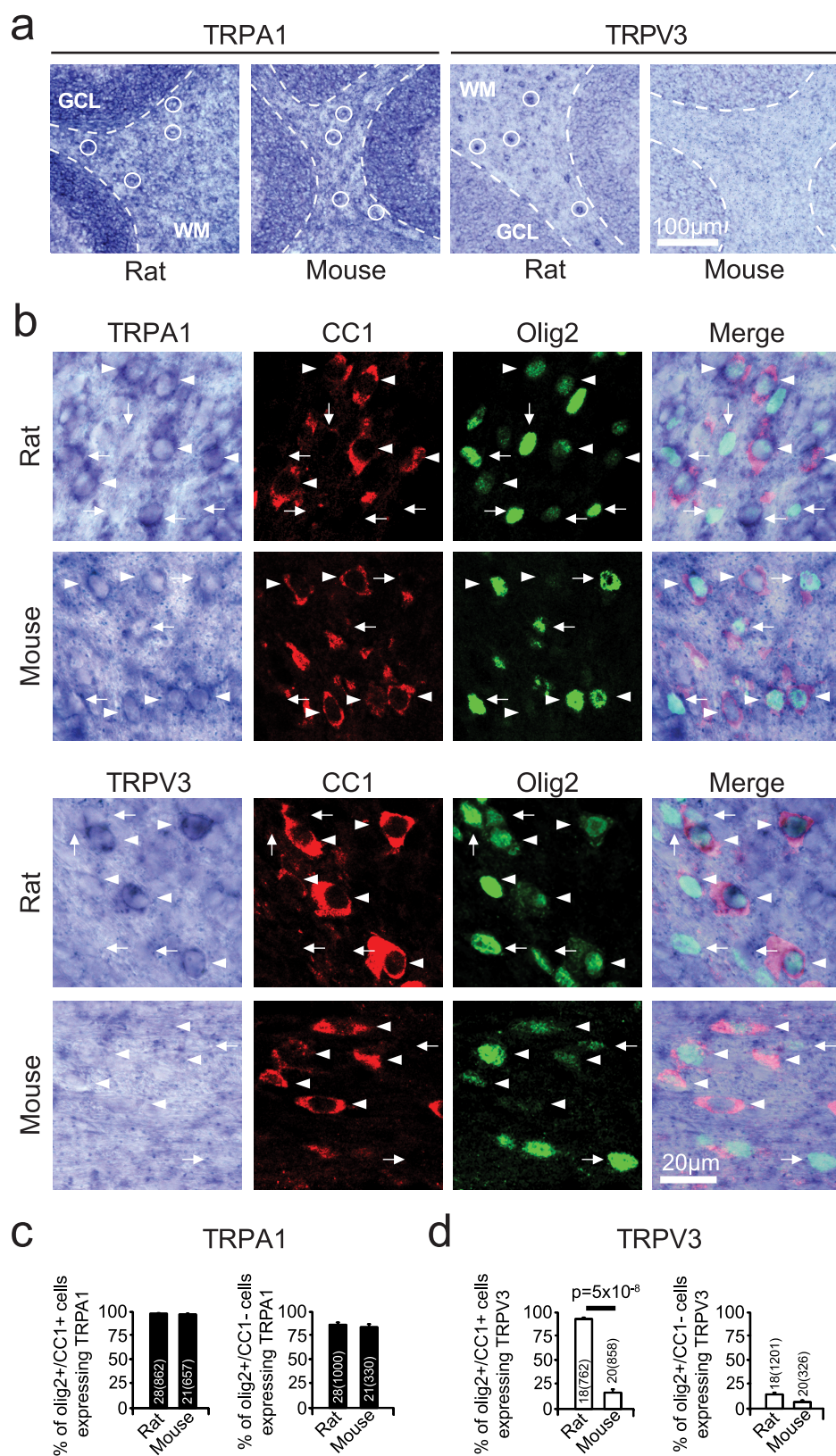
Extended Data Figure 3 | NMDA-evoked ion concentration changes in neurons differ from those in oligodendrocytes. a, Specimen records of rat cerebellar granule cell membrane current and background-subtracted fluorescent dye ratio (R , see Methods) when measuring $[Ca^{2+}]_i$ with Fura-2, $[Na^+]_i$ with SBFI, and $[K^+]_i$ with PBFI, when $100\mu M$ NMDA was

applied. The rise of $[K^+]_i$ seen reflects K^+ entry: $[K^+]_{pipette}$ was 32.5 mM , so $E_K > -60\text{ mV}$ for $[K^+]_o > 3.3\text{ mM}$. **b,** Mean peak fluorescence change normalized to evoked current (number of cells on bars; P value from Mann–Whitney test). Oligodendrocyte data for comparison are shown in Fig. 2. Error bars are s.e.m.



Extended Data Figure 4 | Comparison of NMDA-, $[K^+]_{\text{bath}}$ - and ischaemia-evoked changes of $[H^+]_i$. **a**, Measurements of changes of ratio (R , see Methods) of background-subtracted BCECF fluorescence in rat oligodendrocytes in response to $100\ \mu\text{M}$ NMDA (5 cells) and raising $[K^+]_{\text{bath}}$ from 2.5 to 5 mM with 0.5 mM internal HEPES (6 cells), and

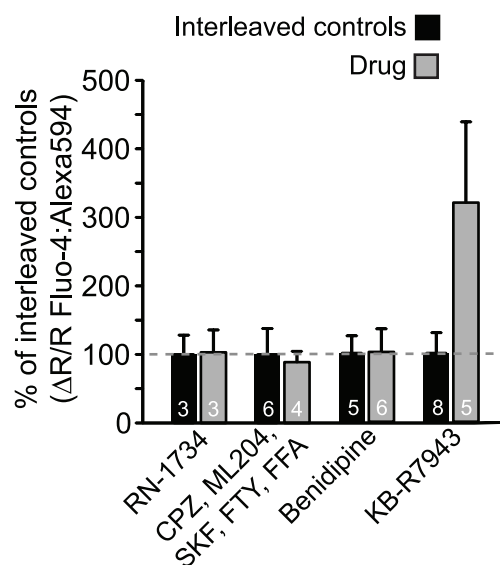
to ischaemia with 0.5 mM and 50 mM internal HEPES (9 and 6 cells, respectively). **b**, **c**, Effect of pH of 10 mM HEPES internal solution on baseline ratio of Fluo-4 to Alexa 594 fluorescence (P value from Mann–Whitney test) (**b**), and change of ratio when $[Ca]$ was increased by 200 nM (**c**). Error bars are s.e.m.



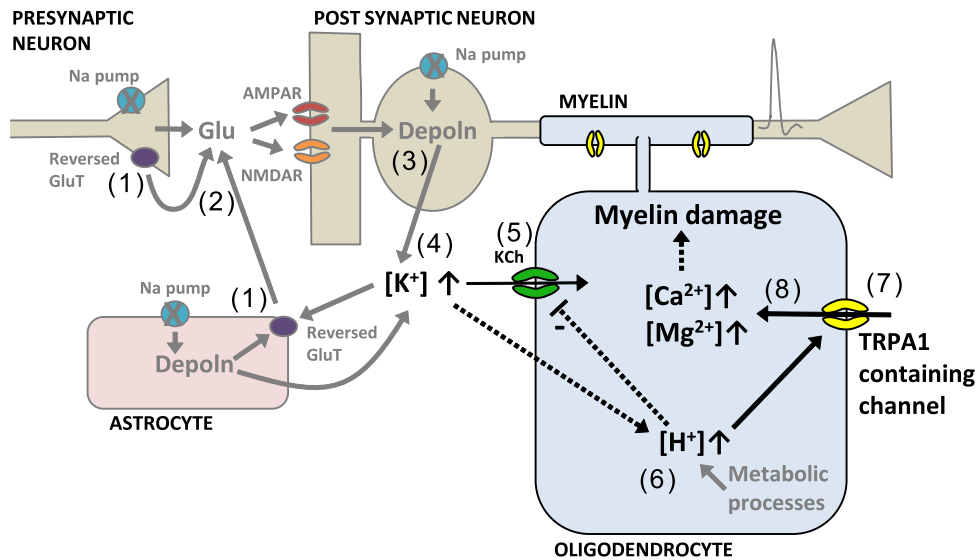
Extended Data Figure 5 | *In situ* hybridization data on TRP channel expression. **a**, *In situ* data for TRPA1 and TRPV3 in the cerebellum of rats and mice show TRPA1 messenger RNA in white matter (WM) cells in rats and mice (with denser expression in the adjacent granule cell layer, GCL), but TRPV3 mRNA only in white matter cells in rats. Specimen cells are labelled with white circles. **b**, Higher magnification views of white matter, combining *in situ* hybridization for TRPA1 and TRPV3 with immunocytochemistry for Olig2 (to label oligodendrocyte lineage

cells) and CC1 (to define myelinating oligodendrocytes). TRPA1 mRNA is present (in rats and mice) and TRPV3 mRNA is present (in rats but not mice) in myelinating oligodendrocytes (Olig2⁺, CC1⁺; arrowheads) and also in some presumed oligodendrocyte precursor cells (Olig2⁺, CC1⁻; arrows). **c**, **d**, Quantification of presence of mRNA for TRPA1 (**c**) and TRPV3 (**d**) in different oligodendrocyte lineage cell classes. Numbers on bars are 'images analysed (cells counted)'. *P* value from Mann–Whitney test. Error bars are s.e.m.

oligodendrocytes



Extended Data Figure 6 | Further evidence for the identity of TRP channels in oligodendrocytes. $[Ca^{2+}]_i$ increase (ratio signal from Fluo-4 and Alexa Fluor 594) in rat oligodendrocyte somata when ischaemia solution was applied with the following drugs present (data normalized to interleaved controls, shown as black bars): RN-1734 (0.5 mM), which blocks TRPV4 and, less well, TRPV1, TRPV3 and TRPM8; a cocktail of blockers inhibiting (see Supplementary Information section on Specificity of drugs acting on TRP channels section) TRPP2, TRPC3, TRPC4, TRPC5, TRPC6, TRPC7, TRPM2, TRPM4, TRPM5, TRPV1, TRPV2, TRPM7, TRPM8 and TRPP1, as well as the store-operated calcium channel component STIM1 and some voltage-gated calcium channels; blocking voltage-gated Ca^{2+} channels with 10 μ M benidipine; or blocking reversed Na/Ca exchange with 10 μ M KB-R7943 mesylate (P values, from Mann-Whitney test and t -test as appropriate, were non-significant ($P > 0.28$)). Error bars are s.e.m.



Extended Data Figure 7 | Schematic of how oligodendrocyte $[Ca^{2+}]_i$ is raised in ischaemia. Run-down of transmembrane ion gradients when Na pump stops (crosses) leads to glutamate transporters (GluT) reversing (1) and releasing glutamate (2). This depolarizes (and causes a neurotoxic $[Ca^{2+}]_i$ rise in) neurons (3) and raises $[K^+]_o$ (4), causing an inward current (5) through oligodendrocyte K^+ channels (KCh). At the

same time, metabolic changes and also the rise of $[K^+]_o$ lead to a rise in oligodendrocyte $[H^+]_i$ (6). This decreases the membrane K^+ conductance (either directly or via TRPA1-containing channels opening) which contributes to the inward current generated, and opens TRPA1-containing channels (7) that let Ca^{2+} and Mg^{2+} into the cell (8). The resulting rise of $[Ca^{2+}]_i$ damages the myelin.

ORIGINAL ARTICLE

Interferon- β therapy specifically reduces pathogenic memory B cells in multiple sclerosis patients by inducing a FAS-mediated apoptosis

Fabiana Rizzo¹, Elena Giacomini¹, Rosella Mechelli², Maria Chiara Buscarinu², Marco Salvetti², Martina Severa^{1,3,4} and Eliana Marina Coccia^{1,3,4}

Growing evidences put B lymphocytes on a central stage in multiple sclerosis (MS) immunopathology. While investigating the effects of interferon- β (IFN- β) therapy, one of the most used first-line disease-modifying drugs for the treatment of relapsing-remitting MS, in circulating B-cell sub-populations, we found a specific and marked decrease of CD27⁺ memory B cells. Interestingly, memory B cells are considered a population with a great disease-driving relevance in MS and resulted to be also target of B-cell depleting therapies. In addition, Epstein–Barr virus (EBV), associated with MS etiopathogenesis, harbors in this cell type and an IFN- β -induced reduction of the memory B-cell compartment, in turn, resulted in a decreased expression of the EBV gene *latent membrane protein 2A* in treated patients. We found that *in vivo* IFN- β therapy specifically and highly induced apoptosis in memory B cells, in accordance with a strong increase of the apoptotic markers Annexin-V and active caspase-3, via a mechanism requiring the FAS-receptor/TACI (transmembrane activator and CAML interactor) signaling. Thus, efficacy of IFN- β therapy in MS may rely not only on its recognized anti-inflammatory activities but also on the specific depletion of memory B cells, considered to be a pathogenic cell subset, reducing their inflammatory impact in target organs.

Immunology and Cell Biology (2016) 94, 886–894; doi:10.1038/icb.2016.55

In multiple sclerosis (MS) research, there is growing acceptance and understanding of B-cell involvement in the immunopathogenesis of this inflammatory demyelinating disease of the central nervous system (CNS), challenging the concept of a unique role for autoreactive T cells.¹

Studies on B-cell-depleting therapies, in particular with the anti-CD20 monoclonal antibody (mAb) Rituximab, boost the knowledge on B-cell functions in MS, unveiling a central role for memory B lymphocytes as the subset with the greatest disease-driving relevance.²

Memory B cells, identified by the expression of the CD27 surface marker,³ were shown to be major producers of proinflammatory cytokines, contributing to the dysregulated inflammatory milieu in MS,⁴ as well as to be effective antigen-presenting cells (APC), thereby soliciting help to autoreactive T cells.⁵ Accordingly, it is noteworthy that Rituximab treatment mediates not only the decrease in the absolute numbers of both peripheral blood and cerebrospinal fluid B cells but also of T cells, suggesting that B cells may contribute to the presence of T cells in MS CNS.⁶

Nonetheless, CD27⁺ memory B cells are the natural reservoirs of life-long latent Epstein–Barr virus (EBV) infection,⁷ an infectious agent that has long been regarded as candidate trigger of autoimmune response in MS.⁸

These observations confirm how targeting the memory B-cell compartment in the treatment of MS might be a successful strategy. Accordingly, in this study we unveil a novel therapeutic action of interferon- β (IFN- β), still one of the most used first-line drugs in the treatment of relapsing–remitting MS (RRMS),⁹ mediating a marked and specific reduction of memory B-cell frequency in the peripheral blood of treated MS patients, similarly as previously found for the B-cell-depleting drug Rituximab.¹ This IFN- β -induced depletion was exerted by the induction of a FAS-receptor (FAS-R)-mediated caspase-3-dependent apoptosis specifically in memory B cells, which in turn resulted also in a significant decrease in the dysregulated EBV *latent membrane protein 2A* (*LMP2A*) gene expression found in the peripheral blood cells of MS patients as compared with paired controls. Thus, IFN- β therapy may mediate B-cell-targeted immunomodulatory activities, and, in turn, induce an EBV-directed anti-viral action exerting a dual effect on MS patients.

¹Department of Infectious, Parasitic and Immune-mediated Disease, Istituto Superiore di Sanità, Rome, Italy and ²Centre for Experimental Neurological Therapies (CENTERS), S Andrea Hospital Site, Sapienza University, Rome, Italy

³Joint senior investigators.

⁴These authors contributed equally to this work.

Correspondence: Dr M Severa or Dr EM Coccia, Department of Infectious, Parasitic and Immune-mediated Disease, Istituto Superiore di Sanità, Viale Regina Elena 299, Rome 00161, Italy.

E-mail: martina.severa@iss.it or eliana.coccia@iss.it

Received 24 March 2016; revised 6 May 2016; accepted 29 May 2016; accepted article preview online 6 June 2016; advance online publication, 5 July 2016

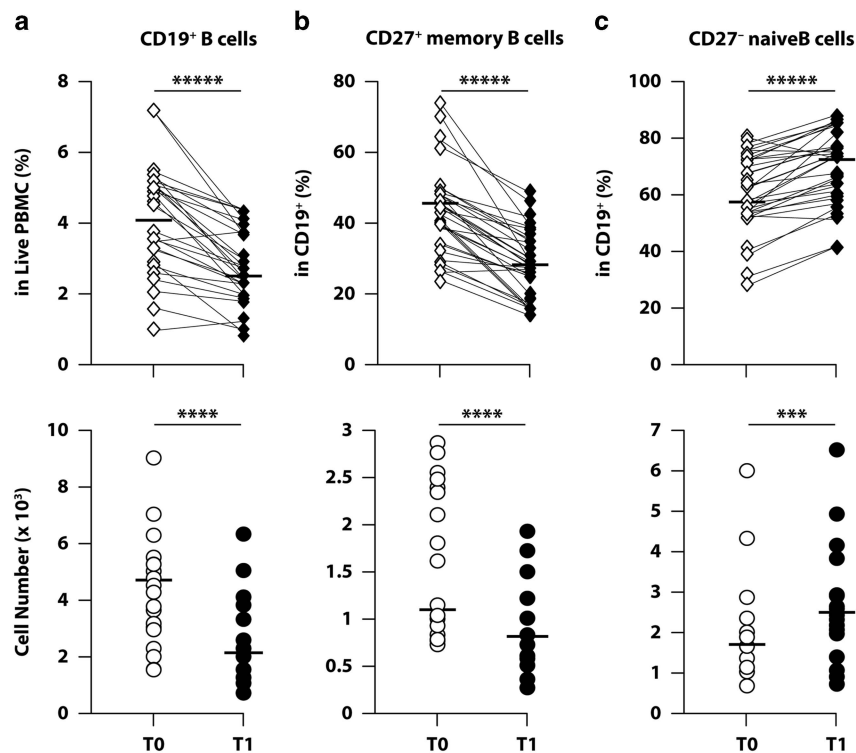


Figure 1 Effects of IFN- β therapy on circulating B-cell populations. Longitudinal studies were performed by flow cytometry on PBMCs isolated from freshly drawn blood of 35 RRMS patients at baseline (no treatment = T0) and after 1 month (T1) of IFN- β therapy. PBMCs were stained with anti-CD19 and CD27 mAbs to gate CD19 $^{+}$ total B cells (a) and discriminate between CD27 $^{+}$ memory (b) or CD27 $^{-}$ naïve (c) B cells. Percentages in gated cells (upper panels) and absolute cell numbers counted in 10^5 gated cells (lower panels) are depicted for each analyzed population, together with median values. *** $P \leq 0.001$; **** $P \leq 0.0001$; ***** $P \leq 0.00001$.

RESULTS

IFN- β treatment specifically reduces the number of circulating memory B cells

Having in mind the powerful immune-modulatory properties of IFN- β , we investigated whether this therapy would regulate the peripheral B-cell subsets in RRMS patients.

Circulating B-cell populations specifically express the lineage marker CD19; however, they can be subdivided in naïve (CD19 $^{+}$ CD27 $^{-}$) or memory (CD19 $^{+}$ CD27 $^{+}$) B-cell subtypes for the selective expression of CD27 surface antigen.¹⁰

First, the frequency of total circulating CD19 $^{+}$ B lymphocytes derived from peripheral blood of MS patients ($n=35$) longitudinally collected before (T0) and after 1 month (T1), from the beginning of IFN- β treatment was assessed by flow cytometry as both percent of live peripheral blood mononuclear cells (PBMCs) and absolute numbers (Figure 1a, upper and lower panel, respectively). This analysis showed that *in vivo* IFN- β administration mediates a marked reduction of whole CD19 $^{+}$ B cells.

Interestingly, the B-cell compartment specifically affected by *in vivo* IFN- β treatment was that of CD19 $^{+}$ CD27 $^{+}$ memory B cells (Figure 1b and Supplementary Figure 1A), with a concomitant increase in the naïve CD19 $^{+}$ CD27 $^{-}$ B-cell counterpart (Figure 1c and Supplementary Figure 1A). This opposite trend was even more pronounced when the frequency of memory and naïve B cells was measured in the same patients following 6 months from the beginning of IFN- β therapy, demonstrating that the effect of this drug was not transitory but rather long-lasting (Supplementary Figure 1B).

In adults, peripheral blood CD27 $^{+}$ memory B cells further divide into Immunoglobulin (Ig)D $^{+}$ CD27 $^{+}$ non-switched memory

and IgD $^{-}$ CD27 $^{+}$ class-switched memory cells.¹⁰ Our results also showed that both non-switched memory and class-switched memory sub-populations of CD27 $^{+}$ memory B cells were markedly decreased upon 1 month of IFN- β treatment; however, non-switched memory B cells were completely lost in treated patients (Supplementary Figure 1C).

The specificity of IFN- β -mediated depletion of memory B cells was further confirmed by the observation that the percentage of circulating CD3 $^{+}$ T cells and CD14 $^{+}$ monocytes remained unchanged (Supplementary Figure 2).

Taken together, these results showed that IFN- β therapy mediates a marked and specific redistribution of the B-cell compartment in MS patients depleting the disease-relevant CD27 $^{+}$ memory B cells.

IFN- β therapy reduces expression of the EBV gene *LMP2A* in PBMCs of RRMS patients

Since IFN- β therapy is targeting circulating CD27 $^{+}$ memory B cells, main reservoirs of EBV infection, we tested whether this drug would also indirectly perturb EBV gene expression levels in MS patients.

EBV establishes a life-long latent infection in resting CD27 $^{+}$ memory B cells, by using the sequential expression of four different transcriptional programs.¹¹ The EBV-encoded *LMP2A*, a functional homolog of the B-cell receptor that conveys survival and antiapoptotic signals to infected cells, is always expressed, even if at various levels, independently of the latency program and allows persistence of latently infected memory B cells in the periphery.¹²

In healthy individuals who experienced EBV infection, circulating virus-harboring B cells are rare (one in 10^5 – 10^6 peripheral B cells).¹³ Thus, we used a very sensitive real-time reverse transcription-PCR

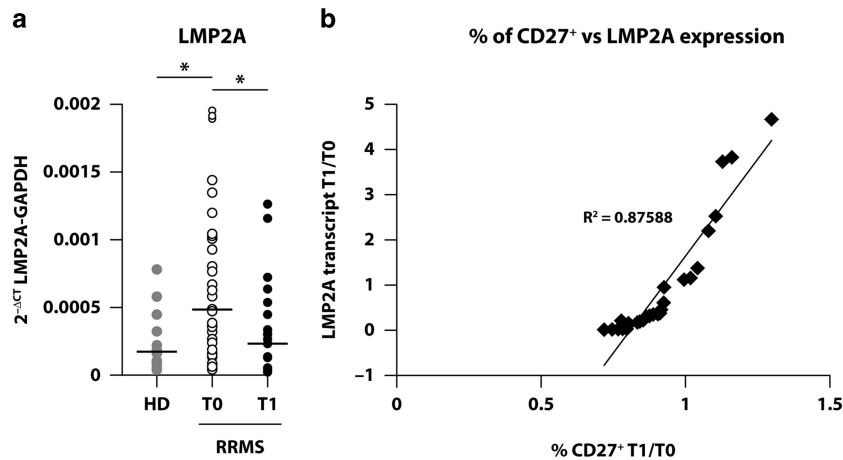


Figure 2 Modulation of EBV-encoded *LMP2A* gene expression by IFN- β therapy. Expression of viral *LMP2A* was analyzed by quantitative real-time reverse transcription-PCR (RT-PCR) in PBMCs of 35 RRMS patients (T0) longitudinally followed upon IFN- β therapy (T1), and in PBMCs of 30 matched HD. (a) *LMP2A* transcription levels were normalized with those obtained for the housekeeping gene *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) by using the formula $2^{-\Delta C_t}$. Normalized values from each single patient and medians are shown. * $P \leq 0.05$. (b) Correlation between the decrease of the *LMP2A* gene expression (fold T1/T0) and the reduced CD27⁺ memory B-cell percentage (fold T1/T0) upon IFN- β therapy is shown.

technique, which includes a specific preamplification step,^{14,15} to measure *LMP2A* transcription in PBMCs of 35 RRMS patients, longitudinally followed before (T0) and 1 month (T1) after the beginning of IFN- β administration, as compared with 30 sex- and age-matched healthy donors (HDs). Our data clearly indicated on the one hand that PBMCs derived from RRMS patients display a significantly higher EBV gene expression than that found in HD, and on the other hand that *in vivo* IFN- β reduces this expression level (Figure 2a). In addition, the decreased *LMP2A* expression observed upon *in vivo* IFN- β treatment positively correlated with the reduction of CD27⁺ memory B-cell frequency found in the same treated patients (Figure 2b).

Thus, we can postulate that IFN- β therapy, by targeting the memory B-cell compartment by means of its well-known immunomodulatory properties, may also exert an anti-viral action in MS patients.

IFN- β therapy does not influence CD27 shedding or memory B-cell homing

We investigated different mechanisms that could have been responsible for the selective reduction of circulating CD27⁺ memory B cells observed in IFN- β -treated MS patients.

Since it was reported that the CD27 transmembrane protein might be cleaved and released in a soluble form (sCD27), whose increased level in sera is assumed to be a marker of abnormalities in B-cell differentiation or activation,^{16,17} at first we measured by enzyme-linked immunosorbent assay (ELISA) sCD27 in the sera of MS patients longitudinally collected at T0 and T1. However, no modification was observed (Figure 3a).

Nonetheless, no increase of the two chemokine receptors CXCR3 and CXCR4 (C-X-C motif chemokine 3 and 4), crucial modulators of B-cell homing and involved in the migration of B cells from the peripheral blood to lymphoid organs or inflamed tissues, respectively,¹⁸ was observed in circulating B cells of paired therapy-naïve or IFN- β -treated MS patients as studied by flow cytometry (Figure 3b).

IFN- β therapy regulates FAS-R and TACI expression on memory B cells of MS patients

We next investigated whether apoptosis might occur in this cell subset, as it is well known that IFN- β may induce controlled cell death

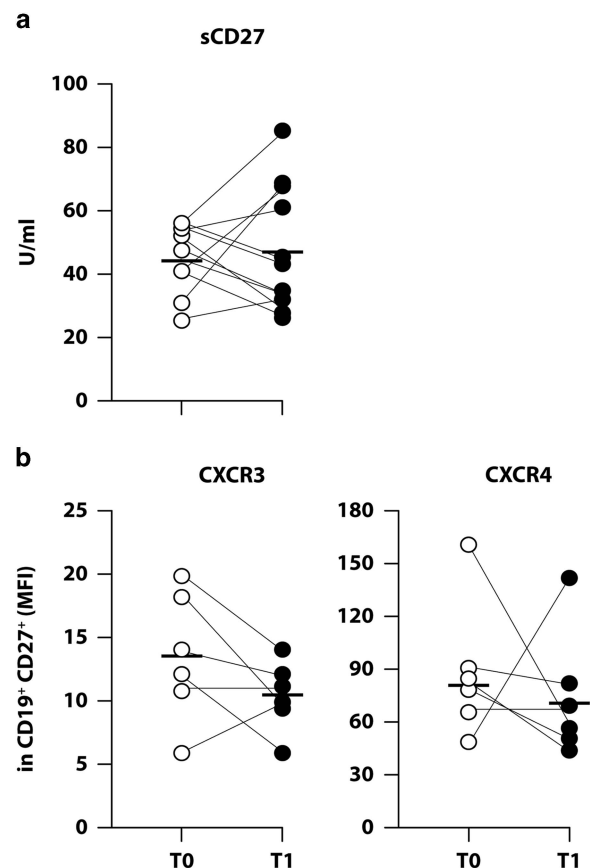


Figure 3 Analysis of possible IFN- β -induced mechanisms mediating the reduction of circulating memory B cells. (a) Levels of sCD27 (sCD27) were analyzed by ELISA in the sera of 11 MS patients longitudinally collected before (T0) and after 1 month of IFN- β treatment (T1). Data from each analyzed patient and median concentration of serum sCD27 are shown. (b) Mean fluorescence intensity (MFI) values (including median) of the chemokine receptors CXCR3 and CXCR4 (C-X-C motif chemokine 3 and 4) in gated CD19⁺ B cells in PBMCs of six RRMS patients, *ex vivo* isolated at T0 and T1, are shown.

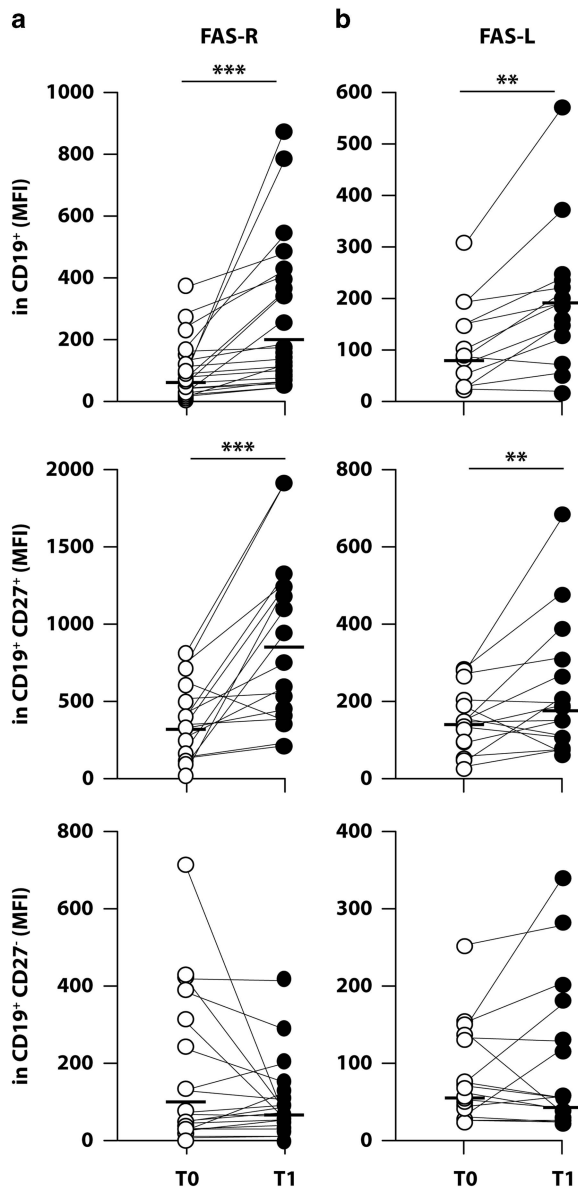


Figure 4 Induction of FAS-R and FAS-L on memory B cells of MS patients in response to IFN- β therapy. FACS analysis of the apoptotic markers FAS-R (a) and FAS-L (b) was performed in PBMCs obtained from 15 RRMS patients before (T0) and after IFN- β treatment (T1). Mean fluorescence intensity (MFI) values obtained for both FAS-R and FAS-L in CD19 $^{+}$ total B cells, CD19 $^{+}$ CD27 $^{+}$ memory and CD19 $^{+}$ CD27 $^{-}$ naive B cells are shown. ** $P \leq 0.01$; *** $P \leq 0.001$.

regulating the transcription of many IFN-inducible genes with proapoptotic functions.¹⁹ The apoptotic marker CD95/FAS-R and its ligand (FAS-L) are among the most important.

Thus, we assumed that IFN- β therapy could mediate the initiation of an apoptotic process in the CD27 $^{+}$ memory B-cell subset of MS patients. Our analysis by flow cytometry on PBMCs of RRMS patients ($n = 15$), isolated before and after 1 month of IFN- β therapy, revealed that the expression of the IFN-inducible apoptotic marker FAS-R was significantly induced on total CD19 $^{+}$ B cells, and, in particular, we observed a much higher level of this receptor on the surface of CD27 $^{+}$ memory B cells with respect to that found in the CD27 $^{-}$ naive

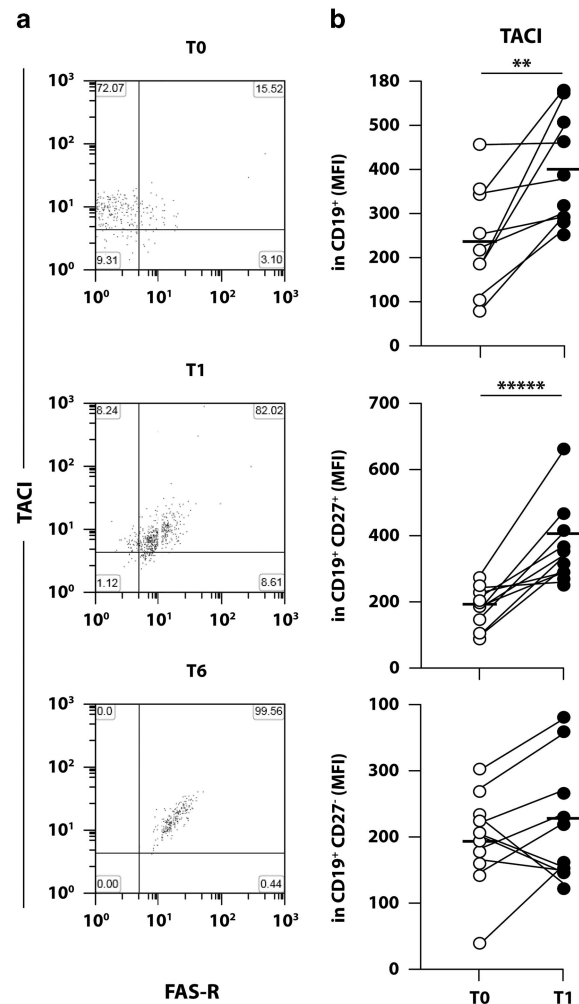


Figure 5 IFN- β -induced increase of TACI surface expression in memory B cells of MS patients. (a) Coexpression of TACI and FAS-R in CD27 $^{+}$ memory B cells was evaluated by FACS analysis in freshly isolated PBMCs of five RRMS patients sampled at baseline (T0), at T1 (1 month) and T6 (6 months) after IFN- β therapy. Representative dot plots and cell frequencies are shown. (b) Levels of TACI expression evaluated in PBMCs of 10 RRMS at T0 and T1 in gated CD19 $^{+}$ total, CD19 $^{+}$ CD27 $^{+}$ memory and CD19 $^{+}$ CD27 $^{-}$ naive B cells are depicted as scatter plots reporting all the obtained mean fluorescence intensity (MFI) values, together with medians. ** $P \leq 0.01$; **** $P \leq 0.00001$.

counterpart (Figure 4a). These results perfectly match the surface expression pattern of FAS-L (Figure 4b).

Thus, a FAS-mediated apoptosis may occur in memory B lymphocytes of MS patients during IFN- β administration and the coexpression of both FAS-R and FAS-L on the same cells may potentiate the induction of cell death in this cell population.

Having in mind transmembrane activator and CAML interactor (TACI) involvement in FAS upregulation and FAS-mediated apoptosis of B cells,^{20,21} we assumed that the expression of this receptor could have been regulated in memory B cells of IFN- β -treated MS patients.

Interestingly, we found that upon 1 month of IFN- β therapy, FAS-R $^{+}$ CD27 $^{+}$ memory B cells also uniformly express TACI, the expression level of which further increased upon 6 months (Figure 5a).

Accordingly, an *ex vivo* longitudinal FACS analysis showed an increased TACI surface expression in CD19⁺ total B cells at T1 as compared with T0, which was specific of CD27⁺ memory B cells and unaffected in the CD27⁻-naïve counterpart (Figure 5b).

FAS-mediated apoptosis is specifically induced in memory B cells in response to IFN- β therapy

To investigate the biological relevance of IFN- β -induced FAS-R/TACI axis on memory B cells, apoptosis was studied in CD27⁺ memory B cells in two complementary ways by cytofluorimetric analysis in PBMCs obtained from longitudinal blood draws of MS patients at T0 and T1. On the one hand, we monitored directly the phosphatidylserine exposure by staining with Annexin-V, and on the other hand we measured the intracellular activation level of caspase-3, the last caspase induced by FAS-R triggering with a central role in the finalization of the apoptotic process.

IFN- β therapy highly and specifically enhances in CD27⁺ memory B cells both Annexin-V expression (Figure 6a) and intracellular caspase-3 activation (Figure 6b). The increase in the two apoptotic markers observed upon *in vivo* IFN- β treatment inversely correlated with the percentage of CD27⁺ memory B cells found in the same treated patients (Figure 6c). Conversely, naïve B-cell subset was not affected (data not shown).

A set of *in vitro* experiments was also performed to verify whether the increased FAS-R expression on memory B cells of MS patients was also functionally active and was directly driving the *in vivo* observed IFN- β -induced apoptosis. Thus, PBMCs isolated from the peripheral blood of MS patients at T0 and T1 were left untreated or treated with an anti-FAS mAb for 20 h (Figure 6d). These cultures were then analyzed by FACS for the expression of both Annexin-V and active caspase-3 in gated CD27⁺ memory B cells showing results in line with the data derived from the *ex vivo* analyzed samples. The *in vivo* IFN- β -treated cells stimulated *in vitro* with the anti-FAS mAb displayed, indeed, an increased expression of both Annexin-V and the active form of caspase-3 in CD27⁺ memory B cells (Figure 6d, left and right panels, respectively).

DISCUSSION

In MS, different therapeutic strategies targeting memory B cells have led to significant reduction in disease activity.^{22,23} The basis for this effect appears to be independent of B-cell Ab production, but rather be related to a decreased production of proinflammatory cytokines or reduced antigen presentation by B cells, activities mediated in particular by the memory B-cell pool.¹

Indeed, in MS patients memory B cells were shown to respond to non-CNS-specific activating stimuli by expressing an exaggerated proinflammatory cytokine response profile.²⁴ Moreover, MS memory B cells were described as potent APC for myelin antigens including myelin-basic protein and myelin-oligodendrocyte glycoprotein eliciting CD4⁺ T-cell proliferation and IFN- γ production.²⁵ Depletion of these cells by way of Rituximab resulted in diminished encephalitogenic T-helper type 1 and 17 proinflammatory T-cell activity.²⁴ Nonetheless, Rituximab therapy, which is also currently in use for EBV-associated B-cell-lymphoproliferative disorders, could effectively deplete EBV-infected memory B cells in the blood of MS patients, displaying a dysregulated EBV response.^{15,26–28}

Moreover, memory B cells express high concentrations of the adhesion molecule very-late antigen-4, which binds to vascular-cell adhesion molecule-1 on endothelial cells.²⁹ The $\alpha 4$ -integrin component of very-late antigen-4 is the target of Natalizumab, another

drug that is able to reduce disease activity and relapse rate in patients with MS.²³ Natalizumab decreases the expression of very-late antigen-4 on circulating immune cells, thereby inhibiting B- and T-cell entry into the CNS, which might explain its beneficial effects in patients with MS.²⁹

In line with these observations, our study reports a novel effect for IFN- β therapy targeting the pathogenic memory B-cell compartment in MS patients.

Although more than two decades have passed since IFN- β was introduced in the management of MS, it remains a valid approach because of its good benefit/risk profile. Its pleiotropic effects appear crucial for MS (an impression substantiated by recent genome-wide association study data^{30–33}) to the point that one may hypothesize that we are still unable to fully exploit the therapeutic potential of this mediator.

It is widely accepted that the beneficial impact of this drug may stem from its anti-inflammatory properties and its broad-range immunomodulatory activities acting on several immune cell types.⁹ Nevertheless, this pleiotropic cytokine not only possesses a wide range of regulatory properties critical for anti-viral immunity but also for impacting cell proliferation and cytoprotection.⁹

Here, we showed that IFN- β therapy mediates a marked and specific reduction of memory B-cell frequency in the peripheral blood of treated MS patients. Similarly to Rituximab that induces cell death of CD20-bearing cells by way of different mechanisms,³⁴ IFN- β -triggered depletion of memory B cells was exerted by the induction of FAS-R-mediated caspase-3-dependent apoptosis, in accordance with a well-known capacity of this cytokine and other type I IFN family members to induce apoptotic cell death regulating the transcription of several IFN-regulated genes with proapoptotic functions.¹⁹

The induction of FAS-R as well as FAS-L specifically on the surface of circulating IFN- β -exposed memory B cells correlated with a strong increase of TACI expression. This receptor is a member of the TNF superfamily and is one of the three receptors for the B-cell-activating factor from the TNF family (BAFF). Upon binding to the BAFF receptor, BAFF induces naïve B-cell survival and maturation; while via the B-cell maturation antigen, BAFF supports the survival of plasma cells in the bone marrow.³⁵

Extensive work was necessary to clarify the more elusive role of TACI, revealing two sides of this receptor: a positive function driving T-cell-independent humoral immune responses³⁶ and a negative one downregulating B-cell activation and expansion by triggering FAS-mediated B-cell apoptosis.²¹ It is also known that an endogenous soluble TACI form exists *in vivo* and that its shedding is due to the proteolytic activity of the metalloproteinase ADAM10, releasing this receptor from the membrane of activated memory B cells.³⁷ Soluble TACI levels are elevated both in the cerebrospinal fluid of MS patients and in the serum of individuals affected by systemic lupus erythematosus correlating with disease activity.³⁷

BAFF exists in two forms, a membrane-bound and a soluble one.³⁸ An *in vitro* study showed that 20 BAFF-soluble trimers associate to form a BAFF 60 mer.³⁹ This structure did not activate the classical BAFF-R transduction, but rather it was shown to drive TACI signaling^{40,41} inducing TRAF-mediated nuclear factor- κ B pathway and, in turn, the transcription of apoptosis-related factors.⁴² The requirement for oligomeric ligands to induce TACI signaling well fits with a model in which both TACI and BAFF are strongly upregulated as in the case of IFN- β -treated MS patients, in which we found that TACI expression is significantly induced in memory B cells and that

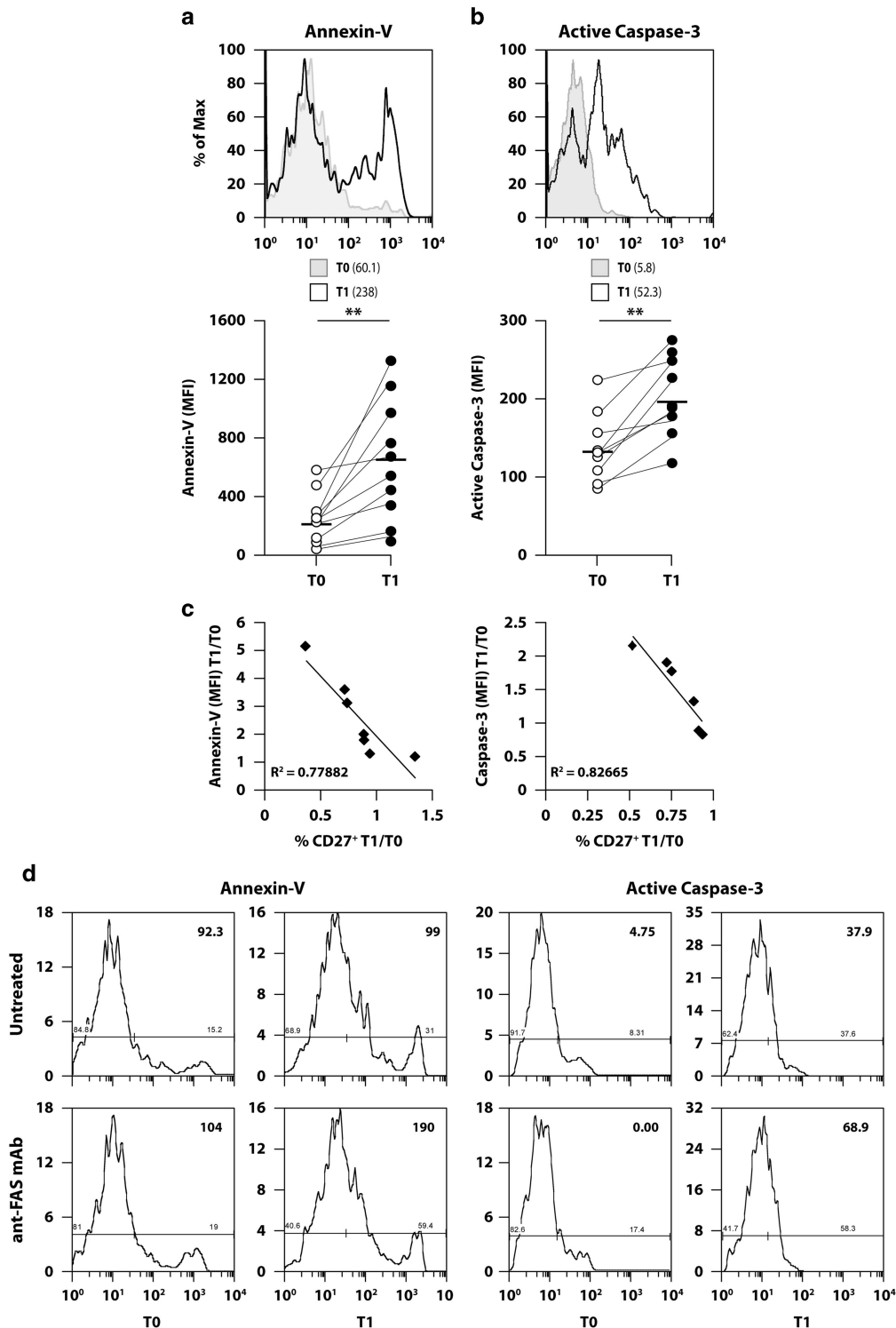


Figure 6 Specific triggering of FAS-mediated apoptosis in memory B cells of MS patients upon IFN- β therapy. (a, b) PBMCs were isolated from 10 MS patients at T0 and T1 and analyzed by flow cytometry for Annexin-V expression (a) and intracellular caspase-3 activation (b) in gated CD19⁺CD27⁺ memory B cells. Data are shown as representative flow plots derived from one MS patient (upper panels) and mean fluorescence intensity (MFI) values obtained from each single analyzed patient (scatter plots, lower panels). ** $P \leq 0.01$. (c) Negative correlation between the IFN- β -induced reduction in CD27⁺ memory B-cell percentage (calculated as fold T1/T0) and the respective induction in Annexin-V expression and caspase-3 activation (calculated as fold T1/T0). (d) PBMCs from 3 RRMS patients were longitudinally isolated before (T0) and after 1 month from the beginning of IFN- β administration (T1). Cells were then cultured for 20 h in the presence of an anti-FAS mAb (100 ng ml⁻¹). Representative flow plots are shown for each culturing condition reporting the percentage of Annexin-V⁺ or active caspase-3⁺ cells and the respective level of expression (MFI values) as evaluated by FACS analysis in CD19⁺CD27⁺ memory B cells.

Table 1 Main demographic and clinical characteristics of MS patients

	Female/male (ratio)	Mean age \pm s.d. (range)	Time since diagnosis in year, median (range)	EDSS, median (range)	Patients with Gd+ lesions on MRI, n (%)
RRMS (n=35)	23/12 (1.9)	36.4 \pm 8 (22–46)	4.8 (0.2–11)	1 (0–2)	13 (37)

Abbreviations: EDSS, expanded disability status scale; Gd, gadolinium; IFN, interferon; MS, multiple sclerosis; MRI, nuclear magnetic resonance imaging. The reported characteristics are referred to patients before the beginning of IFN- β therapy.

BAFF is increased in both sera and peripheral blood cells as compared with therapy-naïve patients.^{43,44}

Our hypothesis is that the strong increase of soluble BAFF observed in IFN- β -treated MS patients may easily support the formation of BAFF 60-mer structures that could, in turn, bind to TACI highly expressed on FAS-R⁺ memory B cells switching from the induction of survival signals to the IFN- β -mediated apoptotic effects specifically in this cell type, without affecting the survival or the maintenance of the long-lived plasma cell pool in the bone marrow.

This hypothesis may also explain the failure of the 'Atacicept in MS' study, a phase II trial conducted to assess safety and efficacy of the fusion protein Atacicept in the treatment of MS.⁴⁵ Atacicept consists of a human recombinant fusion protein of the extracellular ligand-binding portion of the human TACI receptor linked to a recombinant Fc portion of human Ig.⁴⁶ In the Atacicept in MS trial, the relapse rate of all patients receiving weekly subcutaneous injections of this drug doubled compared with placebo controls, leading to early termination of the study.⁴⁵ Targeting of BAFF signal by Atacicept disrupted B-cell development and survival without greatly affecting B-cell counts while substantially reducing the serum levels of all the Ig subtypes,⁴⁵ conversely to what was observed in the case of Rituximab.² A probably more important issue of treatment with this drug was also the type of B cells that might be preferentially depleted. As BAFF is also believed to be involved in the differentiation of a B-cell subset with regulatory properties, regulatory B cells (B regs),⁴⁷ known to have immune-regulatory functions and to constraint T-cell-mediated autoimmune processes,⁴⁸ BAFF targeting by Atacicept might disturb the fine-tuned balance of conventional and B reg in favor of the conventional—more pathogenic—ones, eventually resulting in increased disease activity, as seen in the Atacicept in MS study.⁴⁵

This well sounds with our data showing that upon IFN- β therapy, the decrease in the CD27⁺ compartment correlates with a concomitant increase in the CD27⁺-naïve counterpart, containing the IL-10-producing B-cell subset with regulatory properties,⁴⁹ likely as a result of a renewal of circulating B cells in the peripheral blood of MS patients (Rizzo F *et al.*, unpublished data), as also recently showed by Schubert *et al.*⁵⁰

Another intriguing piece of data of our work is that showing an increase in the expression of the EBV-encoded gene *LMP2A* in the peripheral blood cells of MS patients as compared with paired controls, which in turn significantly decreased the gene expression following IFN- β therapy. The level of the measured *LMP2A* transcription correlated with the frequency of CD27⁺ memory B cells found in the same patients, indicating that IFN- β -mediated targeting of the memory B-cell pool may also associate with an anti-viral activity of this drug in RRMS patients.

Indeed, a selective reduction of peripheral memory B cells may, in turn, result in a decreased migration to the CNS of pathogenic and also possibly EBV-infected B cells, thus affecting disease exacerbation and relapse rate in MS patients.

It is worthy of note that a proportion of the analyzed patients (20–30%) behaved differently than the majority of patients in terms of

CD27⁺ memory B-cell reduction and expression of apoptotic markers in response to IFN- β treatment. One might hypothesize that the IFN- β -induced decrement of this cell subset could predict the responsiveness to this therapy for each single patient. However, further studies and a higher number of patients are needed to clarify this intriguing aspect.

Our study, therefore, makes a contribution in the new emerging theory of a possible usage of anti-viral therapies in autoimmune diseases,⁵¹ and also further characterizes the multitude of targets and pleiotropic effects that IFN- β therapy may have in the treatment of MS. Thus, having in mind IFN- β well-known safety and tolerability profile, we believe that new schedules of administration and/or 'smart' combination therapies may be tested to reduce IFN- β -associated adverse events and increase efficacy.

METHODS

Patients

Thirty-five patients with definite RRMS (age 36.4 \pm 8 years (mean \pm s.d.)) according to the revised McDonald's criteria⁵² were enrolled at S Andrea Hospital (Rome, Italy), following the standard procedures of the outpatients' service of the MS Center. Patients were 66% women and 34% men, respecting the female-to-male ratio of MS, which in Europe varies from 1.5 to 3. Ethics Committee of S Andrea Hospital approved the study (CE 204/10) and all the subjects involved gave written informed consent. The main demographic and clinical characteristics of recruited patients are shown in Table 1. To be included in the study, patients had to be free of immunosuppressive therapies for at least 3 months. Magnetic resonance imaging was performed for each patient within 30 days from sampling.

All the 35 enrolled patients were longitudinally studied right before (T0) and 1 month (T1), and in some cases (n=5) 6 months (T6), after the beginning of IFN- β treatment (recombinant IFN- β 1b in the formulation of Betaferon (Bayer, Lervikusen, Germany) or recombinant IFN- β 1a in the formulation of Rebif 22 or 44 (Merck, Darmstadt, Germany)). We chose to study 1 month after the beginning of therapy to make sure to analyze a time window in which anti-IFN-neutralizing Abs were not produced yet, and thus could not mask or alter patients' responsiveness.

Thirty sex- and age-matched HD (35.4 \pm 9.44 years; 20 females/10 males; ratio = 2) were also analyzed in some of the experiments performed.

PBMC isolation, *in vitro* culture and stimulation

Peripheral blood (40–50 ml) was collected in EDTA-coated tubes from MS patients and processed within 16 h from sampling.

PBMCs were obtained as described previously.^{43,53} Briefly, PBMCs were isolated by density gradient centrifugation using Lympholyte-H (Cedarlane, Hornby, ON, Canada; no. CL5020). Where indicated, PBMCs were resuspended in culture medium consisting of RPMI-1640 (Lonza-BioWhittaker, Verviers, Belgium; no. BE12-167 F) supplemented with 10% fetal bovine serum (Lonza-BioWhittaker; no. DE14-801 F), in the presence of penicillin and streptomycin (Lonza-BioWhittaker; no. DE17-602E) and complemented with L-glutamine (Lonza-BioWhittaker; no. BE17-605E).

Where indicated, PBMCs isolated from peripheral blood of MS patients collected at T0 and T1 were stimulated for 20 h with 100 ng ml⁻¹ of an anti-FAS mAb (Euromedex, Munddsheim, France; no. 05-201).

Flow cytometry analysis

Approximately 5×10^5 PBMCs were aliquoted into appropriate tubes in phosphate-buffered saline (Lonza-BioWhittaker; no. BE17-516 F) containing 2% fetal bovine serum, washed and stained with mAbs, as specified by the manufacturers and incubated at +4 °C for 30 min. Following incubation, the cells were washed in phosphate-buffered saline and fixed in 300 μ l of phosphate-buffered saline containing 2% formaldehyde (Prolabo, Dublin, Ireland; no. 20909290). Stained cells were then evaluated on a FACSCanto (BD Bioscience, San Jose, CA, USA) and phenotypic analysis was performed using Flow Jo software (Tree Star Inc., Ashland, OR, USA).

In particular, to characterize within the lymphocyte gate the different sub-populations of B cells present within the peripheral blood, PBMCs were *ex vivo* isolated from MS patients at T0, T1 and T6 from the beginning of IFN- β therapy, and examined for the expression of specific surface markers characteristic of B-cell subtypes. In particular, CD19 mAb (BD Pharmingen, San Diego, CA, USA; no. 345790) was used to define total circulating B lymphocytes, whereas CD27 mAb (Beckman Coulter; no. B09983), together with IgD mAb (BD Pharmingen; no. 555778), allowed the discrimination among CD19⁺CD27⁻IgD⁺-naïve B cells, non-switched CD19⁺CD27⁺IgD⁺ and class-switched CD19⁺CD27⁺IgD⁻ memory B cells.¹⁰

In addition, to determine the respective percentage of CD3⁺ T cells, CD14⁺ monocytes and CD19⁺ B cells, the CD19-FITC/CD3-PE/CD14-APC mix was used.

To monitor the surface expression of the apoptotic markers CD95/FAS-R, CD178/FAS-L and TACI both in memory and naïve B cells, a mix containing CD19-FITC, CD27-APC added to CD95-PE, CD178-PE (Miltenyi, Bergisch Gladbach, Germany; no. 130096456) or TACI-PE (eBiosciences, San Diego, CA, USA; no. 129217-42) was prepared. Furthermore, to study the coexpression of CD95 and TACI in memory B cells, a mix containing CD27-APC/CD19-FITC was added to TACI-PE and CD95-PacificBlue (BD Pharmingen; no. 562648).

Similarly, mAbs directed to both C-X-C motif chemokine 3 and 4 (BD Pharmingen; nos. 557185 and 555974) were used to monitor expression of the chemokine receptors on total CD19⁺ B cells.

To discriminate live cells from cells undergoing apoptosis, PBMCs were also stained for Annexin-V-FITC (Abcam, Cambridge, UK; no. AB14082), which has a high affinity for the membrane-associated phospholipid phosphatidylserine exposed along the apoptotic process.

Intracellular expression of the active form of caspase-3 was also analyzed by flow cytometry in different B-cell subtypes. Briefly, the cells were stained initially for B-cell surface markers as described above, and then expression of active caspase-3 was evaluated by intracellular staining with an anti-active caspase-3 mAb (BD Pharmingen; no. 561011) by using the BD Cytofix/Cytoperm Fixation/Permeabilization Kit (BD Pharmingen; no. 555028) according to the protocol provided.

The different IgG1 and/or IgG2a isotype negative control Abs were added as needed.

ELISA

Sera were harvested from 10 MS patients at T0 and T1 from the beginning of IFN- β administration and stored at -80 °C. Instant ELISA Kit for sCD27 was purchased from Bender MedSystems (Burlingame, CA, USA; no. BMS286INST). The shewed values showed represent the medians \pm confidence interval of the receptor concentration detected in the sera.

RNA purification and real-time reverse transcription-PCR

DNase-I-treated total RNA was purified from untreated ($n=50$) or IFN- β -treated ($n=35$) PBMCs derived from MS patients or HD ($n=30$) by using the RNeasy Mini Kit (Qiagen Inc., Valencia, CA, USA; no. 74104). The quantity of the extracted RNA was evaluated by Nanodrop spectrophotometer (Nanodrop2000; Thermo Scientific, EuroClone, Waltham, MA, USA).

Reverse transcription of total RNA (500 ng) was primed with oligo (dT) and random primers by the murine leukemia virus reverse transcriptase (Invitrogen, Life Technologies, Carlsbad, CA, USA; no. 28025013). The obtained cDNA was purified using the QIAquick PCR Purification Kit (Qiagen Inc.; no. 28106).

To measure EBV-encoded *LMP2A* expression, the PreAMP Master Mix Kit (Applied Biosystems, Life Technologies, Foster City, CA, USA; no. 4384266) was used to enrich for viral gene transcripts as described previously.¹⁵

Quantitative real-time reverse transcription-PCR assays for the housekeeping gene *GAPDH* and *LMP2A* were conducted at least in duplicate by using the LightCycler Fast Start DNA SYBR Green I Master Mix (Roche Diagnostics, Mannheim, Germany; no. 800-620-627) in the presence of 3 mM MgCl₂ on a LightCycler Instrument (Roche Diagnostics).

Average sample values were normalized by calculating the relative quantity of each mRNA to that of *GAPDH* by using the formula $2^{-\Delta C_t}$, where ΔC_t represents the difference in cycle threshold between target and housekeeping mRNAs.

The primer pairs used in this study were as described previously.¹⁵

Statistical analysis

Results were expressed as median \pm confidence interval. Comparisons between HD and patients were analyzed by the Mann-Whitney *U*-test. The sample size was chosen to ensure that the margin of error was sufficiently small to have a confidence interval of 95%.

Comparisons between patients at different time points of therapy were analyzed by the Wilcoxon's matched-pairs signed-rank test. Correlations were analyzed by the Spearman's rank coefficient test. In all tests, a *P*-value of <0.05 was considered statistically significant. In the figures, star scale was assigned as follows: **P* \leq 0.05; ***P* \leq 0.01; ****P* \leq 0.001; *****P* \leq 0.0001 and ******P* \leq 0.00001.

CONFLICT OF INTEREST

MS received lecture fees from Biogen-Dompé, research support from Bayer-Schering, Biogen-Dompé, Merck-Serono and Sanofi-Aventis.

ACKNOWLEDGEMENTS

This work was supported by the Italian Multiple Sclerosis Foundation projects no. 2013/R/9 (to EMC) and no. 2014/R/19 (to MS). We are thankful to Eugenio Morassi (Division Service for Data Management, Documentation, Library and Publishing Activities, Istituto Superiore di Sanità, Rome, Italy) for preparing drawings.

- 1 von Budingen HC, Bar-Or A, Zamvil SS. B cells in multiple sclerosis: connecting the dots. *Curr Opin Immunol* 2011; **23**: 713–720.
- 2 Barun B, Bar-Or A. Treatment of multiple sclerosis with anti-CD20 antibodies. *Clin Immunol* 2012; **142**: 31–37.
- 3 Klein U, Rajewsky K, Kuppers R. Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. *J Exp Med* 1998; **188**: 1679–1689.
- 4 Duddy ME, Alter A, Bar-Or A. Distinct profiles of human B cell effector cytokines: a role in immune regulation? *J Immunol* 2004; **172**: 3422–3427.
- 5 Bhargava P, Calabresi PA. Novel therapies for memory cells in autoimmune diseases. *Clin Exp Immunol* 2015; **180**: 353–360.
- 6 Monson NL, Cravens PD, Frohman EM, Hawker K, Racke MK. Effect of rituximab on the peripheral blood and cerebrospinal fluid B cells in patients with primary progressive multiple sclerosis. *Arch Neurol* 2005; **62**: 258–264.
- 7 Young LS, Rickinson AB. Epstein-Barr virus: 40 years on. *Nat Rev Cancer* 2004; **4**: 757–768.
- 8 Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Ann Neurol* 2007; **61**: 288–299.
- 9 Severa M, Rizzo F, Giacomini E, Salvetti M, Coccia EM. IFN-beta and multiple sclerosis: cross-talking of immune cells and integration of immunoregulatory networks. *Cytokine Growth Factor Rev* 2015; **26**: 229–239.
- 10 Anderson SM, Tomayko MM, Shlomchik MJ. Intrinsic properties of human and murine memory B cells. *Immunol Rev* 2006; **211**: 280–294.
- 11 Thorley-Lawson DA. Epstein-Barr virus: exploiting the immune system. *Nat Rev Immunol* 2001; **1**: 75–82.
- 12 Caldwell RG, Wilson JB, Anderson SJ, Longnecker R. Epstein-Barr virus LMP2A drives B cell development and survival in the absence of normal B cell receptor signals. *Immunity* 1998; **9**: 405–411.
- 13 Wagner HJ, Bein G, Bitsch A, Kirchner H. Detection and quantification of latently infected B lymphocytes in Epstein-Barr virus-seropositive, healthy individuals by polymerase chain reaction. *J Clin Microbiol* 1992; **30**: 2826–2829.

- 14 Angelini DF, Serafini B, Piras E, Severa M, Coccia EM, Rosicarelli B *et al*. Increased CD8⁺ T cell response to Epstein-Barr virus lytic antigens in the active phase of multiple sclerosis. *PLoS Pathog* 2013; **9**: e1003220.
- 15 Serafini B, Severa M, Columba-Cabezas S, Rosicarelli B, Veroni C, Chiappetta G *et al*. Epstein-Barr virus latent infection and BAFF expression in B cells in the multiple sclerosis brain: implications for viral persistence and intrathecal B-cell activation. *J Neuropathol Exp Neurol* 2010; **69**: 677–693.
- 16 Bohnhorst JO, Bjorgan MB, Thoen JE, Jonsson R, Natvig JB, Thompson KM. Abnormal B cell differentiation in primary Sjogren's syndrome results in a depressed percentage of circulating memory B cells and elevated levels of soluble CD27 that correlate with Serum IgG concentration. *Clin Immunol* 2002; **103**: 79–88.
- 17 De Roos AJ, Mirick DK, Edlefsen KL, LaCroix AZ, Kopecky KJ, Madeleine MM *et al*. Markers of B-cell activation in relation to risk of non-Hodgkin lymphoma. *Cancer Res* 2012; **72**: 4733–4743.
- 18 Muehlinghaus G, Cigliano L, Huehn S, Peddinghaus A, Leyendeckers H, Hauser AE *et al*. Regulation of CXCR3 and CXCR4 expression during terminal differentiation of memory B cells into plasma cells. *Blood* 2005; **105**: 3965–3971.
- 19 Chawla-Sarkar M, Lindner DJ, Liu YF, Williams BR, Sen GC, Silverman RH *et al*. Apoptosis and interferons: role of interferon-stimulated genes as mediators of apoptosis. *Apoptosis* 2003; **8**: 237–249.
- 20 Mackay F, Schneider P. Cracking the BAFF code. *Nat Rev Immunol* 2009; **9**: 491–502.
- 21 Figgett WA, Fairfax K, Vincent FB, Le Page MA, Katik I, Deliyanti D *et al*. The TACI receptor regulates T-cell-independent marginal zone B cell responses through innate activation-induced cell death. *Immunity* 2013; **39**: 573–583.
- 22 Hauser SL, Waubant E, Arnold DL, Vollmer T, Antel J, Fox RJ *et al*. B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. *N Engl J Med* 2008; **358**: 676–688.
- 23 Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH *et al*. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med* 2006; **354**: 899–910.
- 24 Bar-Or A, Fawaz L, Fan B, Darlington PJ, Rieger A, Ghorayeb C *et al*. Abnormal B-cell cytokine responses a trigger of T-cell-mediated disease in MS? *Ann Neurol* 2010; **67**: 452–461.
- 25 Harp CT, Ireland S, Davis LS, Remington G, Cassidy B, Cravens PD *et al*. Memory B cells from a subset of treatment-naïve relapsing-remitting multiple sclerosis patients elicit CD4⁽⁺⁾ T-cell proliferation and IFN- γ production in response to myelin basic protein and myelin oligodendrocyte glycoprotein. *Eur J Immunol* 2010; **40**: 2942–2956.
- 26 Serafini B, Rosicarelli B, Franciotta D, Magliozzi R, Reynolds R, Cinque P *et al*. Dysregulated Epstein-Barr virus infection in the multiple sclerosis brain. *J Exp Med* 2007; **204**: 2899–2912.
- 27 Farrell RA, Antony D, Wall GR, Clark DA, Fisman L, Swanton J *et al*. Humoral immune response to EBV in multiple sclerosis is associated with disease activity on MRI. *Neurology* 2009; **73**: 32–38.
- 28 Jilek S, Schluep M, Meylan P, Vingerhoets F, Guignard L, Monney A *et al*. Strong EBV-specific CD8⁺ T-cell response in patients with early multiple sclerosis. *Brain* 2008; **131**(Part 7): 1712–1721.
- 29 Niino M, Bodner C, Simard ML, Alatab S, Gano D, Kim HJ *et al*. Natalizumab effects on immune cell responses in multiple sclerosis. *Ann Neurol* 2006; **59**: 748–754.
- 30 Croze E, Yamaguchi KD, Knappertz V, Reder AT, Salamon H. Interferon-beta-1b-induced short- and long-term signatures of treatment activity in multiple sclerosis. *Pharmacogenom J* 2013; **13**: 443–451.
- 31 Esposito F, Sorosina M, Ottoboni L, Lim ET, Replogle JM, Raj T *et al*. A pharmacogenetic study implicates SLC9a9 in multiple sclerosis disease activity. *Ann Neurol* 2015; **78**: 115–127.
- 32 International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, Hellenthal G, Pirinen M, Spencer CC *et al*. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011; **476**: 214–219.
- 33 International Multiple Sclerosis Genetics C, Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kempainen A *et al*. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* 2013; **45**: 1353–1360.
- 34 Chan AC, Carter PJ. Therapeutic antibodies for autoimmunity and inflammation. *Nat Rev Immunol* 2010; **10**: 301–316.
- 35 Figgett WA, Vincent FB, Saulep-Easton D, Mackay F. Roles of ligands from the TNF superfamily in B cell development, function, and regulation. *Semin Immunol* 2014; **26**: 191–202.
- 36 von Bulow GU, van Deursen JM, Bram RJ. Regulation of the T-independent humoral response by TACI. *Immunity* 2001; **14**: 573–582.
- 37 Hoffmann FS, Kuhn PH, Laurent SA, Hauck SM, Berer K, Wendlinger SA *et al*. The immunoregulator soluble TACI is released by ADAM10 and reflects B cell activation in autoimmunity. *J Immunol* 2015; **194**: 542–552.
- 38 Abu-Rish EY, Amrani Y, Browning MJ. Toll-like receptor 9 activation induces expression of membrane-bound B-cell activating factor (BAFF) on human B cells and leads to increased proliferation in response to both soluble and membrane-bound BAFF. *Rheumatology (Oxford)* 2013; **52**: 1190–1201.
- 39 Liu Y, Xu L, Opalka N, Kappler J, Shu HB, Zhang G. Crystal structure of sTALL-1 reveals a virus-like assembly of TNF family ligands. *Cell* 2002; **108**: 383–394.
- 40 Bossen C, Cachero TG, Tardivel A, Ingold K, Willen L, Dobles M *et al*. TACI, unlike BAFF-R, is solely activated by oligomeric BAFF and APRIL to support survival of activated B cells and plasmablasts. *Blood* 2008; **111**: 1004–1012.
- 41 Cachero TG, Schwartz IM, Qian F, Day ES, Bossen C, Ingold K *et al*. Formation of virus-like clusters is an intrinsic property of the tumor necrosis factor family member BAFF (B cell activating factor). *Biochemistry* 2006; **45**: 2006–2013.
- 42 Zhang Y, Li J, Zhang YM, Zhang XM, Tao J. Effect of TACI signaling on humoral immunity and autoimmune diseases. *J Immunol Res* 2015; **2015**: 247426.
- 43 Giacomini E, Severa M, Rizzo F, Mechelli R, Annibaldi V, Ristori G *et al*. IFN-beta therapy modulates B-cell and monocyte crosstalk via TLR7 in multiple sclerosis patients. *Eur J Immunol* 2013; **43**: 1963–1972.
- 44 Krumbholz M, Faber H, Steinmeyer F, Hoffmann LA, Kumpfel T, Pellkofer H *et al*. Interferon-beta increases BAFF levels in multiple sclerosis: implications for B cell autoimmunity. *Brain* 2008; **131**(Part 6): 1455–1463.
- 45 Kappos L, Hartung HP, Freedman MS, Boyko A, Radu EW, Mikol DD *et al*. Atacicept in multiple sclerosis (ATAMS): a randomised, placebo-controlled, double-blind, phase 2 trial. *Lancet Neurol* 2014; **13**: 353–363.
- 46 Hartung HP, Kieseier BC. Atacicept: targeting B cells in multiple sclerosis. *Ther Adv Neurol Disord* 2010; **3**: 205–216.
- 47 Yang M, Sun L, Wang S, Ko KH, Xu H, Zheng BJ *et al*. Novel function of B cell-activating factor in the induction of IL-10-producing regulatory B cells. *J Immunol* 2010; **184**: 3321–3325.
- 48 Mauri C, Blair PA. Regulatory B cells in autoimmunity: developments and controversies. *Nat Rev Rheumatol* 2010; **6**: 636–643.
- 49 Blair PA, Norena LY, Flores-Borja F, Rawlings DJ, Isenberg DA, Ehrenstein MR *et al*. CD19⁽⁺⁾CD24^(hi)CD38^(hi) B cells exhibit regulatory capacity in healthy individuals but are functionally impaired in systemic lupus erythematosus patients. *Immunity* 2010; **32**: 129–140.
- 50 Schubert RD, Hu Y, Kumar G, Szeto S, Abraham P, Winderl J *et al*. IFN-beta treatment requires B cells for efficacy in neuroautoimmunity. *J Immunol* 2015; **194**: 2110–2116.
- 51 Dreyfus DH. Autoimmune disease: a role for new anti-viral therapies? *Autoimmun Rev* 2011; **11**: 88–97.
- 52 Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L *et al*. Diagnostic criteria for multiple sclerosis: 2005 revisions to the 'McDonald Criteria'. *Ann Neurol* 2005; **58**: 840–846.
- 53 Severa M, Rizzo F, Giacomini E, Annibaldi V, Gafa V, Romano S *et al*. IFN-beta therapy regulates TLR7-mediated response in plasmacytoid dendritic cells of multiple sclerosis patients influencing an anti-inflammatory status. *J Interferon Cytokine Res* 2015; **35**: 668–681.

The Supplementary Information that accompanies this paper is available on the Immunology and Cell Biology website (<http://www.nature.com/icb>)

SCIENTIFIC REPORTS

OPEN

Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls

Received: 19 April 2016

Accepted: 03 June 2016

Published: 27 June 2016

Jun Chen¹, Nicholas Chia^{2,3}, Krishna R. Kalari¹, Janet Z. Yao², Martina Novotna^{4,5}, M. Mateo Paz Soldan⁴, David H. Luckey⁶, Eric V. Marietta⁷, Patricio R. Jeraldo², Xianfeng Chen¹, Brian G. Weinshenker⁴, Moses Rodriguez^{4,6}, Orhun H. Kantarci⁴, Heidi Nelson², Joseph A. Murray^{6,7} & Ashutosh K. Mangalam^{6,8}

Multiple sclerosis (MS) is an immune-mediated disease, the etiology of which involves both genetic and environmental factors. The exact nature of the environmental factors responsible for predisposition to MS remains elusive; however, it's hypothesized that gastrointestinal microbiota might play an important role in pathogenesis of MS. Therefore, this study was designed to investigate whether gut microbiota are altered in MS by comparing the fecal microbiota in relapsing remitting MS (RRMS) (n = 31) patients to that of age- and gender-matched healthy controls (n = 36). Phylotype profiles of the gut microbial populations were generated using hypervariable tag sequencing of the V3–V5 region of the 16S ribosomal RNA gene. Detailed fecal microbiome analyses revealed that MS patients had distinct microbial community profile compared to healthy controls. We observed an increased abundance of *Pseudomonas*, *Mycoplana*, *Haemophilus*, *Blautia*, and *Dorea* genera in MS patients, whereas control group showed increased abundance of *Parabacteroides*, *Adlercreutzia* and *Prevotella* genera. Thus our study is consistent with the hypothesis that MS patients have gut microbial dysbiosis and further study is needed to better understand their role in the etiopathogenesis of MS.

Multiple sclerosis (MS) is a pro-inflammatory demyelinating disease of the central nervous system (CNS)¹. Most MS patients (~85%) present with a relapsing-remitting MS (RRMS) disease course characterized by clearly defined attacks or relapses, followed by a variable degree of recovery. The etiology of MS is complex and poorly understood. Both genetic and environmental factors play a role^{2,3}, and recent evidence suggests that gut microbiota is one of the key environmental factors. According to the “hygiene hypothesis”, reduced exposure to infections in childhood may increase the risk of allergic and autoimmune diseases^{4,5}. Supporting this argument, Western societies report an increased incidence of diseases with an autoimmune/allergic component, including MS. Increased constipation and fecal incontinence⁶ and increased gut permeability⁷ in MS patients, and increased occurrence of inflammatory bowel diseases (IBD) in MS patients and their families^{8,9}, suggest an important gut–CNS connection. Interestingly, gut bacteria can also influence the blood brain barrier integrity¹⁰. These studies implicate that gut microbiota may potentially be operational in predisposition to or modification of the disease course of MS. Therefore we hypothesized that RRMS patients have gut microbial dysbiosis compared to healthy controls.

To test our hypothesis, we analyze the fecal microbiota composition in patients in the active or remission phases of RRMS and compare it to that of healthy controls. We show that RRMS patients have a distinct microbial community profile compared to healthy controls.

¹Division of Biomedical Statistics and Informatics—Department of Health Sciences Research Mayo Clinic, 200 1st ST SW, Rochester, MN -55905, USA. ²Department of Surgical Research Mayo Clinic, 200 1st ST SW, Rochester, MN -55905, USA. ³Department of Biophysics Mayo Clinic, 200 1st ST SW, Rochester, MN -55905, USA. ⁴Mayo Clinic Center for Multiple Sclerosis and CNS Demyelinating Diseases, Department of Neurology, Mayo Clinic College of Medicine, 200 1st ST SW, Rochester, MN-55905, USA. ⁵International Clinical Research Center, St. Anne's University Hospital Brno, Pekařská 53, 656 91 Brno, Czech Republic. ⁶Department of Immunology Mayo Clinic, 200 1st ST SW, Rochester, MN -55905, USA. ⁷Department of Gastroenterology, Mayo Clinic, Rochester, MN-55905, USA. ⁸Department of Pathology, 25 S Grand Ave, 1080-ML, University of Iowa, Iowa City, IA-52242, USA. Correspondence and requests for materials should be addressed to A.K.M. (email: ashutosh-mangalam@uiowa.edu)

	RRMS			Control
	Active	Remission	Total	
N	12	19	31	36
Sex (M/F)	3/9	7/12	10/21	14/36
Age (SD)	39.3 (10.6)	45.2 (10.2)	42.9 (10.6)	40.3 (7.3)
Age of onset	32.6 (7.9)	37.2 (11.5)	35.4 (10.4)	–
BMI	32.7 (7.4)	26.2 (5.5)	28.0 (6.3)	27.8 (4.7)
Disease severity				–
EDSS				
<3	7	12	19	
3–5	3	0	3	
>6	0	2	2	
ND	3	5	8	
Therapy				
IFNb	7	7	14	
Copax	0	1	1	
Tysabri	1	4	5	
None	4	7	11	

Table 1. Clinical and demographic features of MS patients and control.

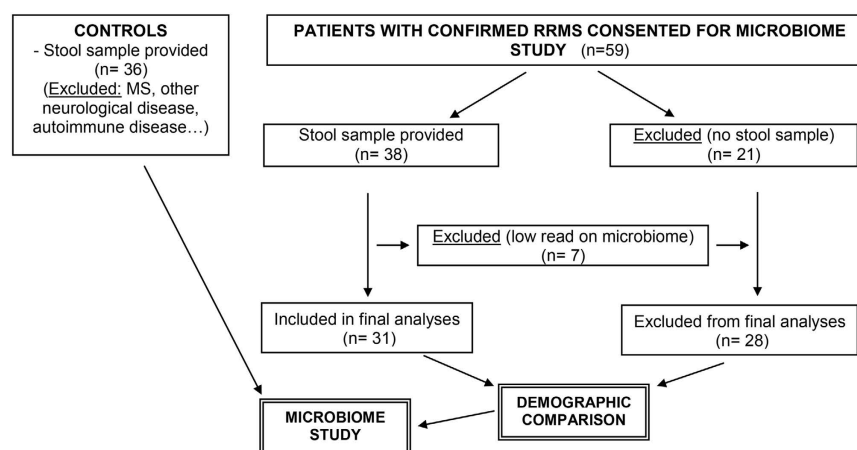


Figure 1. Flowchart explaining RRMS patients and controls recruited for the study with exclusion and inclusion criteria.

Results

Gut Microbiota of RRMS Patients Differs from Healthy controls. To determine if MS is associated with a change in microbial diversity, fecal samples from 31 RRMS patients were sequenced (Table 1 and Fig. 1) and analyzed using IM_TORNADO¹¹. The median sequencing length obtained was 58,272 bases (range 2,658–894,587). After removing singletons, these sequences were clustered into OTUs based on 97% sequence similarity. Taxonomic classification revealed a typical Western diet diversity profile comprising Firmicutes (58.6%), Bacteroidetes (40.4%), Proteobacteria (0.7%), Actinobacteria (0.1%), and a tail of rare bacterial phyla (0.2%) (Supplemental Figure S1). The overall species richness of the gut microbiota of the RRMS patients was not significantly different from that of the healthy controls ($P = 0.73$ for observed OTU number, Supplemental Figure S2). However, when RRMS patients were divided into patients with active disease and those in remission, there was a trend towards lower species richness in patients with active disease compared to healthy controls ($P = 0.1$). The microbiota in patients in the remission phase exhibited species richness similar to the healthy controls (Fig. 2A). The Shannon diversity index, which considers both the species richness and evenness, also identified a similar decreasing trend in species richness for patients in the active disease state ($P = 0.2$, Fig. 2B).

Bray-Curtis distance-based community analysis revealed that the microbiota structure differed significantly between RRMS patients in both remission and active state and healthy controls ($P < 0.001$, PERMANOVA test, Fig. 2C). Interestingly, the structure of the active state microbiota was also different from that of the remission state ($P = 0.05$, PERMANOVA, Fig. 1D), with the remission state microbiota being more similar to that of the healthy controls ($P = 0.06$, 1000 permutations; Fig. 2D). The remission group appeared to be composed of two subgroups, one that was similar to the active state, and one that was similar to the healthy controls (data not shown). The heterogeneity of the remission group resulted in more variability in the entire RRMS group, as

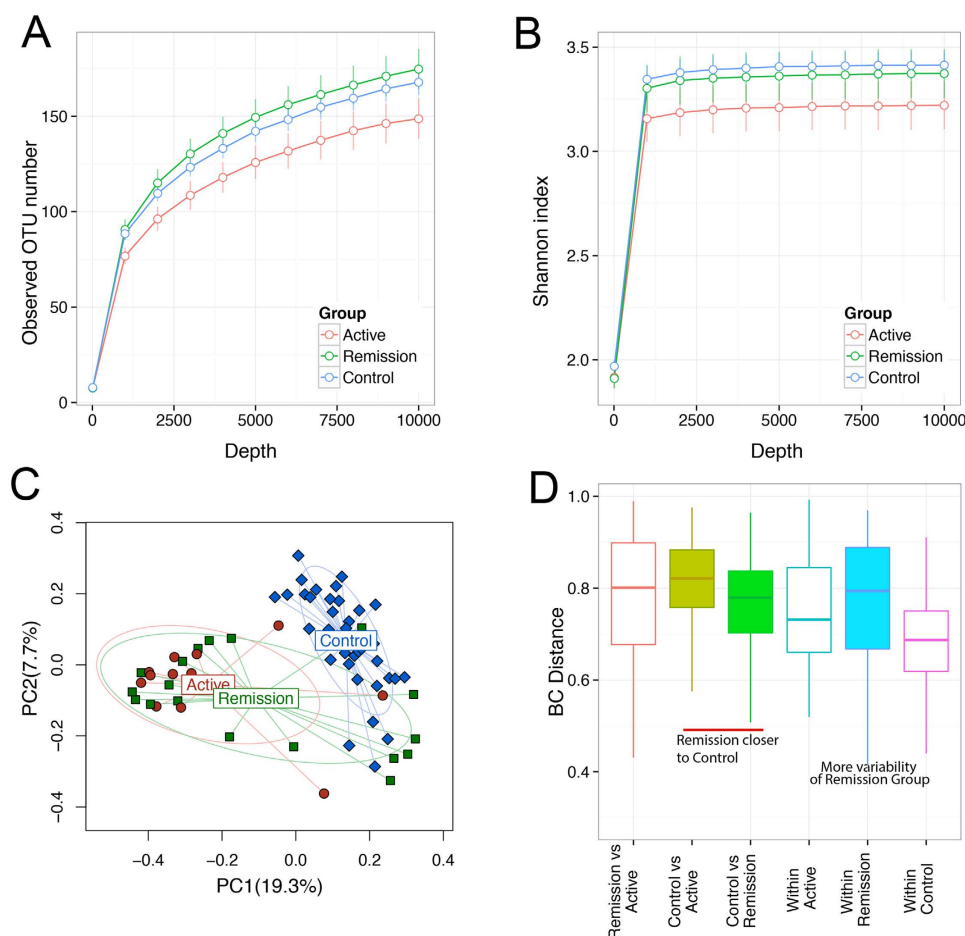


Figure 2. Gut microbiota of MS patients differs from healthy controls. (A,B) Rarefaction curves comparing the species richness (observed OTU number) and overall diversity (Shannon index) between controls, remission RRMS and active RRMS. The microbiota of active disease has lower diversity. (C) Principal coordinate plot based on Bray-Curtis distance matrix. The first two coordinates are plotted with the percentage of variability explained indicated on the axis. Each point represents a sample with colors representing different states. The control samples are significantly different from MS samples. The remission phase samples show greater heterogeneity with subsets resembling control or active disease samples. The ellipses do not represent any statistical significance but rather serve a visual guide to group differences. (D) Boxplots comparing Bray-Curtis distances between various groups. The three horizontal lines of the box represent the first, second (median) and third quartiles respectively with the whisk extending to 1.5 inter-quartile range (IQR).

well as in the remission group, compared to the control group as revealed by a distance-based homogeneity test ($P < 0.001$, PERMDISP test; Fig. 2D). We did not observe any effect of treatment status, smoking, or vitamin D supplementation on composition of microbiota (data not shown).

Abundance of Certain Bacteria Is Associated with RRMS. We used a conservative Wilcoxon rank-sum test to perform differential abundance analyses at the phylum, family, and genus levels, confining the analyses to the taxa with prevalence $>10\%$ and maximum proportion >0.002 . At a false discovery rate of 5%, we identified 35 differentially abundant taxa (Fig. 3A, Table 2). Within the phylum Actinobacteria, the genera *Adlercreutzia* and *Collinsella* were less abundant in RRMS patients compared to controls. Among Bacteroidetes, *Pedobacter* and *Flavobacterium* genera showed higher abundance, whereas *Parabacteroides* showed lower abundance in MS patients compared to healthy controls. Certain Firmicutes genera such as *Blautia* and *Dorea* were enriched in RRMS patients, whereas others belonging to families Erysipelotrichaceae, Lachnospiraceae, Veillonellaceae and genera *Lactobacillus* and *Coprobacillus* showed lower abundance in RRMS patients compared to healthy controls. Among Proteobacteria, *Pseudomonas* and *Mycoplana* were more abundant in RRMS patients, whereas *Haemophilus* was more abundant in healthy controls (Fig. 3B). Using the relative abundance of these differential abundant taxa, we could separate the RRMS microbiota from the control microbiota (Fig. 3C).

Eleven taxa (Table 2, highlighted in bold) remained significant even when stringent Bonferroni correction (P -value cutoff = 0.0005) was used for addressing multiple testing. Using a two log fold changes as a cut-off, we observed that RRMS patients showed enrichment of *Pseudomonas* (Proteobacteria), *Pedobacter* (Bacteroidetes), *Blautia* (Firmicutes), *Dorea* (Firmicutes), and *Mycoplana* (Proteobacteria); whereas healthy controls showed

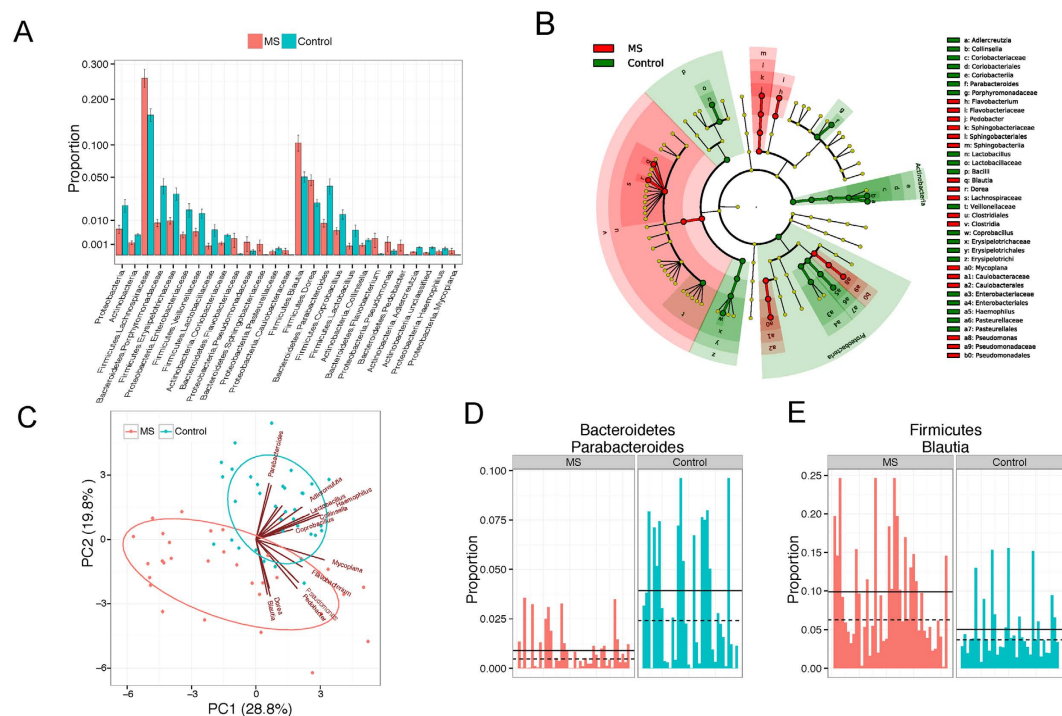


Figure 3. Microbial signatures of the gut microbiota of MS patients. (A) Barplots comparing the abundances of differentially abundant taxa between MS and control. These “signature” taxa are selected by Wilcoxon rank-sum tests and a false discovery rate of 5%. Error bars represent standard errors. Phylum, family and genus-level taxa are plotted. (B) The “signature” taxa are highlighted on the phylogenetic tree (cladogram) using GraPhlAn with red and green color indicating increase and decreases of abundance in the MS patients. (C) Biplots based on principal component analysis of the abundances of the “signature” genera. The first two components are plotted with the percentage of variability explained indicated. Each point represents a sample. The contribution of each genus to the principal components is indicated by the angle and length of the blue line. (D,E) Barplots comparing the relative abundances of *Parabacteroides* and *Blautia* between MS and control. Each bar represents the relative abundance of a sample.

enrichment of *Adlercreutzia* (Actinobacteria), *Collinsella* (Actinobacteria), *Lactobacillus* (Firmicutes) and *Parabacteroides* (Bacteroidetes). Figure 3D shows the abundance data from representative gut microbes increased in control (*Parabacteroides*) or RRMS patients (*Blautia*). Thus, these data indicate that RRMS patients had differential abundance of certain gut microbiota compared to controls.

Predictive Model Using RF. We next assessed the predictive power of the gut microbiota using RF, which predicts disease status based on an ensemble of decision trees. We used RF to build a predictive model based on the gut microbiota profile using the genus-level abundance data as the input. Based on 500 bootstrap samples, we achieved a mean classification error of 0.08, compared to 0.49 based on random guess, meaning we always predicted the class label (RRMS status) of the test sample to be the class label of the majority class seen in the training set ($P < 2.2 \times 10^{-16}$, Friedman Rank Sum test, Fig. 4A). RRMS samples differ substantially in genus-level abundance from the control samples. The relative importance of each genus in the predictive model was assessed using mean decreasing accuracy and Gini coefficient. We used the Boruta algorithm to select significant genera, and 18 genera were confirmed for their importance in prediction (Fig. 4B). This prediction model confirmed that certain genera such as *Adlercreutzia*, *Pedobacter*, *Pseudomonas*, *Coprobacillus*, *Dorea*, *Flavobacterium*, *Parabacteroides*, *Mycoplana*, *Haemophilus*, *Blautia*, and *Collinsella* were predictive of the disease state. Identification of the same genera in the earlier abundance analysis (Fig. 3A) indicated the robustness of both analyses. Hierarchical clustering based on the abundance profile of these confirmed genera showed that the MS samples generally clustered together (Fig. 4C).

We repeated the analysis using the abundance data of 291 OTUs (Fig. 5). Even lower classification error of 0.04 was achieved based on these bootstrap samples. The Boruta algorithm confirmed 25 OTUs, with the majority of them belonging to the same genera identified above (Fig. 5B,C). The new genera identified by the analysis were *Prevotella*, *Clostridium*, and *Erwinia*. OTU belonging to Certain *Prevotella* species was observed more frequently in healthy controls compared to RRMS patients (Supplemental Figure S3). The high concordance between the results obtained from two methods indicated that the RF prediction model was a powerful tool for selecting microbes that were associated with the RRMS.

Functional Analysis. We used PICRUSt to infer the functional content of the microbiota based on closed-reference OTU picking. Twenty COG functional categories were tested. At an FDR of 0.05, we identified

	P value	Q value	MS Mean	Control Mean	Log2 fold change	MS prevalence	Control prevalence
Phylum*							
Actinobacteria	4.09E-06	3.27E-05	1.32E-03	3.46E-03	−1.39	31	36
Proteobacteria	3.60E-03	1.44E-02	5.63E-03	2.03E-02	−1.85	29	36
Family*							
Actinobacteria; Coriobacteriaceae	2.13E-06	2.49E-05	1.22E-03	3.34E-03	−1.46	31	36
Bacteroidetes; Flavobacteriaceae	1.76E-03	7.70E-03	2.26E-03	1.41E-05	7.33	25	14
Bacteroidetes; Porphyromonadaceae	4.85E-04	3.40E-03	8.60E-03	3.95E-02	−2.20	30	36
Bacteroidetes; Sphingobacteriaceae	6.69E-08	2.34E-06	1.03E-03	1.89E-06	9.09	21	1
Firmicutes; Erysipelotrichaceae	6.90E-04	3.63E-03	9.77E-03	3.10E-02	−1.67	31	36
Firmicutes; Lachnospiraceae	2.67E-03	1.04E-02	2.58E-01	1.61E-01	0.68	31	36
Firmicutes; Lactobacillaceae	5.38E-03	1.71E-02	7.24E-04	5.30E-03	−2.87	24	33
Firmicutes; Veillonellaceae	7.25E-04	3.63E-03	4.54E-03	1.44E-02	−1.67	31	36
Proteobacteria; Caulobacteraceae	3.58E-04	3.13E-03	2.03E-04	1.38E-06	7.20	15	3
Proteobacteria; Enterobacteriaceae	9.75E-03	2.84E-02	3.44E-03	1.69E-02	−2.29	27	35
Proteobacteria; Pasteurellaceae	4.35E-03	1.52E-02	1.07E-04	3.67E-04	−1.78	6	18
Proteobacteria; Pseudomonadaceae	6.96E-07	1.22E-05	1.44E-03	1.51E-04	3.25	23	3
Genus*							
Actinobacteria; Adlercreutzia	1.39E-07	4.16E-06	8.99E-05	5.16E-04	−2.52	25	34
Actinobacteria; Collinsella	6.87E-03	3.17E-02	8.69E-04	1.87E-03	−1.11	26	30
Actinobacteria; unclassified	1.80E-03	1.34E-02	4.57E-05	4.92E-04	−3.43	19	24
Bacteroidetes; Flavobacterium	1.76E-03	1.34E-02	2.26E-03	1.41E-05	7.33	25	14
Bacteroidetes; Parabacteroides	7.25E-04	7.25E-03	8.51E-03	3.93E-02	−2.21	30	36
Bacteroidetes; Pedobacter	6.69E-08	4.01E-06	1.03E-03	1.89E-06	9.09	21	1
Firmicutes; Blautia	4.38E-04	5.25E-03	1.04E-01	5.03E-02	1.04	31	36
Firmicutes; Coprobacillus	2.24E-03	1.34E-02	5.01E-03	1.38E-02	−1.46	31	36
Firmicutes; Dorea	2.05E-03	1.34E-02	4.66E-02	2.26E-02	1.05	31	36
Firmicutes; Lactobacillus	5.81E-03	2.90E-02	7.23E-04	5.24E-03	−2.86	24	33
Proteobacteria; Haemophilus	4.35E-03	2.37E-02	1.07E-04	3.67E-04	−1.78	6	18
Proteobacteria; Mycoplasma	3.58E-04	5.25E-03	2.03E-04	1.38E-06	7.20	15	3
Proteobacteria; Pseudomonas	6.96E-07	1.39E-05	1.44E-03	1.51E-04	3.25	23	3

Table 2. Differentially abundant taxa between MS and control samples at phylum, family and genus-level.

*-An FDR of 0.05 is used to identify these taxa.

10 differentially abundant COGs (Fig. 6A), Supplemental Table S4). Four COG categories, including signal transduction mechanism, lipid transport and metabolism, intracellular trafficking, and defense mechanisms, exhibited the most significant differences (Fig. 6B), Supplemental Table S4). In contrast to the taxonomic abundance, the functional abundance exhibited much less variability, with fold changes ranging narrowly between 1% and 6%. However, even under such small effect sizes we were still able to detect differentially abundant COGs because of even smaller within-group variability. We also compared 104 KEGG pathways. At an FDR of 0.05, we identified 34 differentially abundant pathways, indicating diverse change in the functions of the RRMS microbiota compared to controls (Fig. 6C). Three KEGG pathways, fatty acid biosynthesis, glycolysis, porphyrin and chlorophyll metabolism, and transporters survived the Bonferroni correction.

Discussion

The present study demonstrates that RRMS patients have a distinct fecal microbiome compared to healthy controls, with certain gut microbes showing decreased or increased abundance in RRMS patients compared to controls. The strength of our study lies in the abundance analyses, as well as utilization of prediction models to identify the bacterial taxa differentially expressed in the disease state. A high concordance rate between abundance and the prediction model is an added advantage of our study. Thus, our data points toward an important role of the gut microbiota in the RRMS patients. These results provide clear evidence that a larger independent replication study focusing on the identified findings will be needed in which MS phase-specific outcome analyses can also be conducted.

In our study, analysis of species richness (α -diversity) showed no difference between total RRMS patients and healthy controls; however, RRMS patients with active disease showed decreased species richness compared to patients in remission and controls. Decreased species richness in RRMS patients with active disease, points towards an important role of gut microbiota in disease exacerbation. Future studies analyzing gut microbiota at various time points might help in a better understanding of the role of gut microbiota in disease exacerbation. We also observed a lower abundance of *Bacteroidetes* in RRMS patients, which is in agreement with an earlier study showing significant decrease in *Clostridium* and *Bacteroidetes* species in RRMS patients¹². In another exploratory study of 15 MS patients, certain taxa such as *Fecalibacterium* were decreased in MS patients compared to healthy

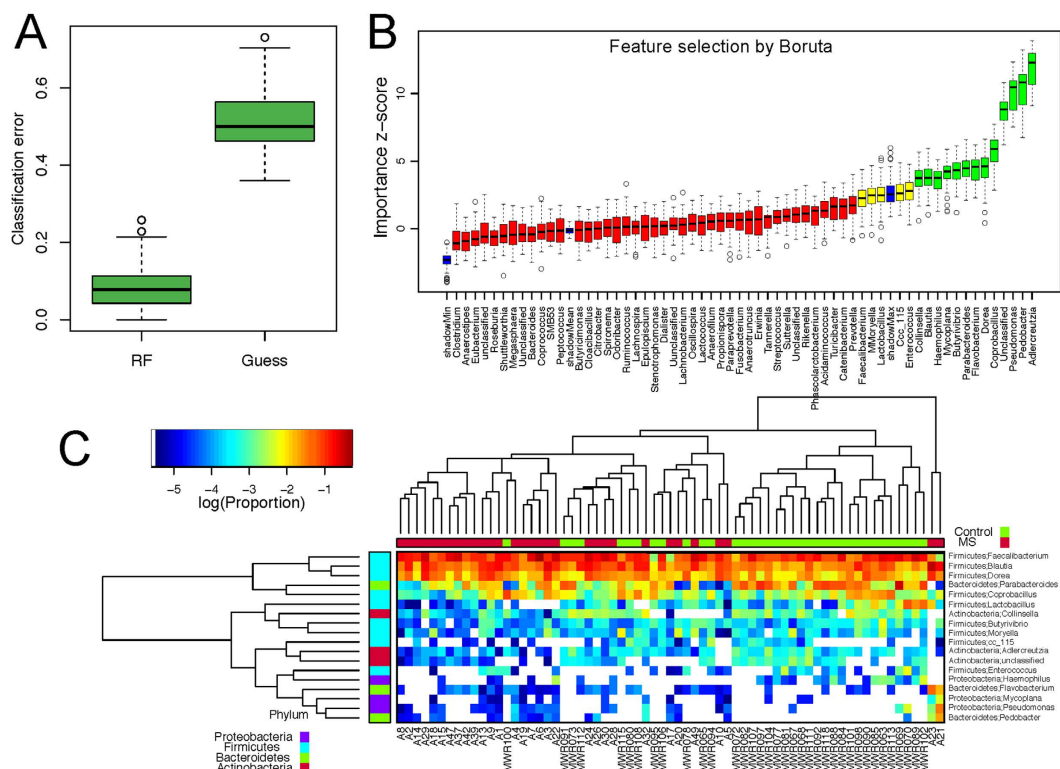


Figure 4. Predictive model based on the genus-level abundance profile using Random Forests (RF).

(A) Comparison of the classification error of the RF trained model to guess, which always predicts the class label based on the majority class in the training data set. The boxplots are based on the results from 500 bootstrap samples. The three horizontal lines of the box represent the first, second (median) and third quartiles respectively with the whisk extending to 1.5 inter-quartile range (IQR). RF achieves significantly lower classification error. (B) Predictive power of individual genera assessed by Boruta feature selection algorithm. Blue boxplots correspond to minimal, average and maximum Z score of shadow genera, which are shuffled version of real genera introduced to RF classifier and act as benchmarks to detect truly predictive genera. Red, yellow and green colors represent rejected, suggestive and confirmed genera by Boruta Selection. (C) Heatmap based on the abundance of Boruta-selected genera. Hierarchical clustering (Euclidean distance, complete linkage) shows that MS samples tend to cluster together.

controls¹³. However we observed similar abundance of *Fecalibacterium* between RRMS and healthy controls. A number of factors such as different 16s primer sets, sequencing techniques, and geographical location might be responsible for the difference in abundance of *Fecalibacterium* between two studies.

Thus our study is in agreement with earlier reports indicating that MS patients had dysbiosis of fecal microbiota in the gut. We did not observe any effect of factors such as smoking, family history, or treatment status on abundance of microbiota; however, we cannot rule out their effect on gut microbiota due to the small sample size in each group. An increased Firmicutes in RRMS with active disease cannot be attributed to obesity alone as a recent meta-analysis of human gut microbiota associated with IBD and obesity had shown that ratio of Firmicutes:Bacteroidetes is not a consistent feature distinguishing lean from obese human microbiota¹⁴.

Our study was not biased by specific pre-determined candidate strata based on a previous gut microbiome analyses. We nevertheless identified several organisms—including *Parabacteroides* and *Prevotella* (Bacteroidetes), *Adlercreutzia* and *Collinsella* (Actinobacteria), and *Erysipelotrichaceae* (Firmicutes)—that were decreased in RRMS compared to healthy controls. Interestingly, *Prevotella*, *Parabacteroides*, and *Adlercreutzia* are associated with the metabolism of phytoestrogens and plant-derived xenoestrogen^{15–17}, whereas *Parabacteroides* and *Erysipelotrichaceae* are involved in bile acid metabolism. Metabolites derived from the metabolism of phytoestrogens (lignan and isoflavone) and bile acids play an important role in maintaining homeostasis at mucosal surfaces through the induction of anti-inflammatory responses¹⁸. Low estrogen states such as menopause and the postpartum period are clearly associated with increased activity in women with MS¹⁹. Further, treatment with estrogens can suppress and/or protect animals from disease in an experimental model of human MS, experimental autoimmune encephalomyelitis (EAE)^{20,21}. When combined with the fact that RRMS patients have reduced levels of bacteria responsible for metabolizing phytoestrogen, this might implicate regulation of estrogen receptor signaling by gut microbiota and/or their metabolites in the etiopathogenesis of RRMS.

Erysipelotrichaceae and *Veillonellaceae* families (Firmicutes) showed lower abundance in RRMS cases compared to controls. *Erysipelotrichaceae* plays an important role in bile acid metabolism²², which can induce anti-inflammatory properties²³. In one report patients with Crohn's disease had decreased

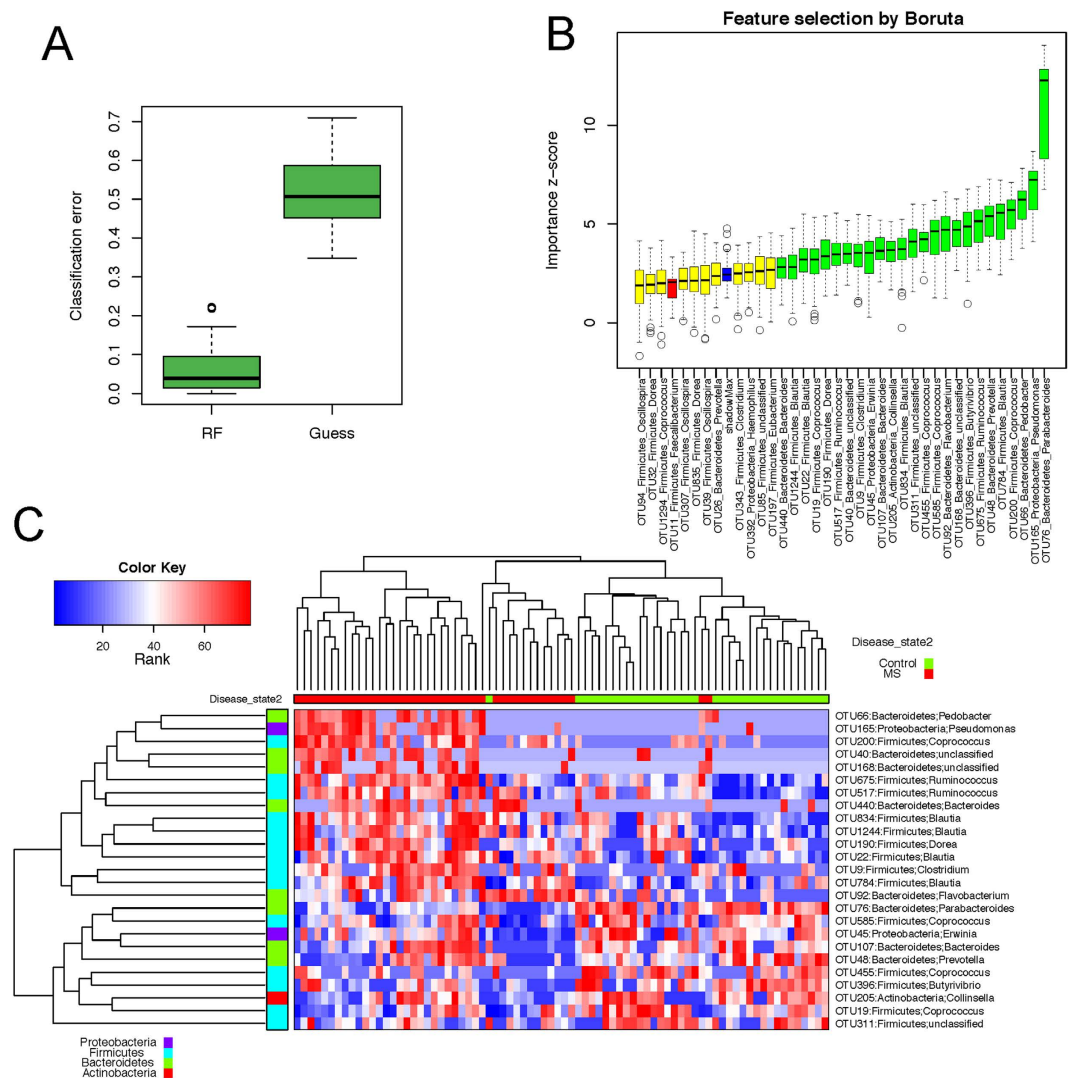


Figure 5. Predictive model based on the OTU level abundance profile using Random Forests (RF). (A) Comparison of the classification error of the RF trained model to guess, which always predicts the class label based on the majority class in the training data set. The boxplots are based on the results from 500 bootstrap samples. The three horizontal lines of the box represent the first, second (median) and third quartile respectively with the whisk extending to 1.5 inter-quartile range (IQR). RF achieves significantly lower classification error. (B) Predictive power of individual genera assessed by Boruta feature selection algorithm. Blue boxplots correspond to minimal, average and maximum Z score of shadow genera, which are shuffled version of real genera introduced to RF classifier and act as benchmarks to detect truly predictive genera. Red, yellow and green colors represent rejected, suggestive and confirmed genera by Boruta Selection. (C) Heatmap based on the abundance Boruta selected genera. Hierarchical clustering (Euclidean distance, complete linkage) shows that MS samples tend to cluster together.

levels of Erysipelotrichaceae in their fecal samples and intestinal biopsies²⁴. Veillonellaceae is closely related to Clostridium, a beneficial commensal that has been shown to induce regulatory T cells²⁵.

RF OTU analysis indicated a decreased abundance of *Prevotella* in MS patients compared to controls. Decreased *Prevotella* levels have been reported in diseases such as type-1-diabetes and autism^{26,27}. In a separate study (manuscript under review) from our lab, we have observed that one of the *Prevotella* species- *Prevotella histicola* can suppress disease in experimental model of MS.

Among Proteobacteria, *Caulobacteraceae*, *Pseudomonas* and *Mycoplana* were more abundant in MS patients, whereas Enterobacteriales were more abundant in the control group. Previous studies have suggested an association between certain Proteobacteria and the pathogenesis of IBD²⁸. Two Firmicutes, *Blautia* and *Dorea*, were more abundant in patients with RRMS compared to controls, however they have been linked both positively²⁹ and negatively³⁰ with inflammatory diseases. Increased *Pedobacter* in microbiome samples has been attributed to contaminated DNA extraction kit reagents³¹. Future studies on large patients' cohort as well as metagenomics studies with help in defining the role of gut microbiota which are abundant in MS patients.

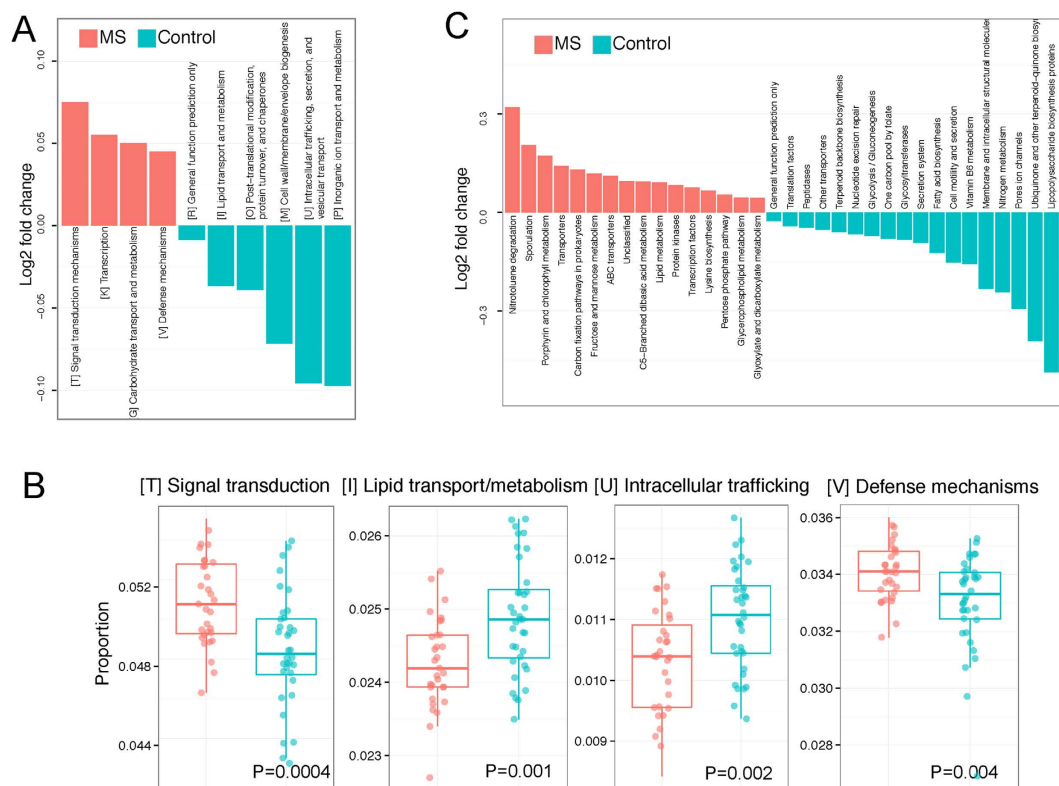


Figure 6. Functional analysis of the MS gut microbiota. PICRUSt was used to infer the functional content of the gut microbiota based on the 16S data. **(A)** Log2 fold change of the abundances of COG categories showing significant difference between MS and control at a false discovery rate of 5%. Red and blue colors indicate increase and decrease in the MS samples. **(B)** Boxplots comparing the abundance distribution of the top 4 differential abundant COG categories. **(C)** Log2 fold change of the abundances of KEGG categories showing significant difference between MS and control at a false discovery rate of 5%.

The functional analysis (PICRUSt) data demonstrated modulation of pathways involved in fatty acid metabolism, defense mechanisms, and glycolysis, which further strengthens our observations from the abundance and RF analyses. We observed a decrease in the abundance of microbes involved in fatty acid metabolism (bile metabolism) and an increase of pathways involved in the defense mechanisms (induction of anti-inflammatory pathways by phytoestrogen and bile acids metabolites). However, the sequencing method alone does have its limitations, as it does not give direct data on the functionally important changes of the microbiota.

The gut/nutrient-centric nature of human evolution suggests an important role for gut-associated microbiota in maintaining homeostasis and human health. Thus, a perturbation of gut microbiota can result in the depletion of certain microbial communities and important metabolites that can lead to impaired immune homeostasis resulting in predisposition to inflammatory disease such as RRMS. Decreased species richness in MS patients with a recent attack compared to those in remission, and modulation of microbes associated with the induction of anti-inflammatory pathways, highlight the importance of the gut microbiome in the etiology of MS. The abundance of microbes involved in the phytoestrogen metabolic pathway is interesting as there is increased female prevalence in MS patients. Based on our observations, gut microbiota may be one of the missing environmental factors responsible for the precipitation of disease in genetically susceptible individuals. Large studies that include time series of samples and methodology to study the functional changes in the intestinal microbiota are needed for the evaluation of the role of gut microbiota in the modulation of immune system response.

Methods

Standard Protocol Approvals, Registrations, and Patient Consents. The study was done in accordance with the guidelines approved by Mayo Clinic Institutional Review Board (IRB). A prior written informed consent was obtained from all the subjects to participate in the study.

Patients and Samples. RRMS patients (n = 62) between 18 and 80 years of age and fulfilling McDonald diagnostic criteria for MS³² with an expanded disability status scale (EDSS) score between 1 and 6 were recruited at the MS Clinic at Mayo Clinic (Rochester, MN). The demographic data for the RRMS patients and healthy controls are summarized in Table 1 and Fig. 1. We also collected data on smoking, vitamin D supplement, and treatment status. Fifty-nine patients with confirmed RRMS consented for microbiome study, 38 of whom provided stool samples. Seven samples were excluded due to low read on sequencing. Thirty-one samples were included for analysis (Fig. 1).

RRMS patients were defined as being in an active disease state if the stool sample was collected within a month of a relapse; otherwise, they were considered to be in remission. Healthy controls ($n = 36$) included an age- and sex-matched cohort with no known disease symptoms. The general exclusion criteria included prior surgeries, including colectomy, ileectomy, and gastrectomy, and excluded minor procedures, such as cholecystectomy, hemorrhoidectomy, and appendectomy. Any patients or controls currently taking antibiotics or probiotic supplements, or having a known history of disease with an autoimmune component such as MS, rheumatoid arthritis, type-1-diabetes, and IBD, were also excluded from the study.

Sample Collection, 16S Amplicon Preparation, Sequencing, and Processing. Subjects were provided with Commode Specimen Collection kits (Fisher Scientific Inc., Pittsburgh, PA, USA) to collect stool. If patients were unable to give the stool sample during their stay at Mayo Clinic, then they were instructed to collect the sample at home and send it through overnight FedEx delivery. Samples were frozen at -70°C within 24 hours of receipt. Microbial DNA was extracted from fecal material of each sample using the MoBio PowerSoil Kit (MoBio Laboratories, Carlsbad, CA, USA) as per the manufacturer's instruction with a bead-beating step. Sequencing of the V3–V5 region of the 16S rRNA was performed as described previously¹¹. The raw 16S data were processed by IM-TORNADO¹¹ to form operational taxonomic units (OTUs) at 97% similarity level.

Statistical Analyses. α -diversity (Observed OTU number and Shannon index) and β -diversity (Bray-Curtis and UniFrac distances) measures were calculated based on the rarefied OTU counts. PERMANOVA was used to test for association between disease status and the overall microbiota composition based on distance matrices ("adonis" function in R package "vegan"). PERMDISP2 was used to test the homogeneity of group dispersions ("betadisper" function in R package "vegan"). Significance was assessed by 1000 permutations. Differential abundance analysis was performed using the Wilcoxon rank-sum test at phylum, family, and genus levels. False discovery rate (FDR) control based on the Benjamini-Hochberg procedure was used to correct for multiple testing. Differentially abundant taxa were highlighted on the tree using the GraPhlAn software (huttenhower.sph.harvard.edu/graphlan).

Predictive Modeling using Random Forests. Random Forests (RF) was used to predict the disease status based on the microbiota profile (genus-level relative abundance data) using default parameters of the R implementation of the algorithm (R package "randomForest"). Bootstrapping ($n = 500$) was used to assess the classification accuracy. The classification performance was compared to random guess, where the class label of a test sample was predicted to be the label of the majority class in the training set, and significance of difference was assessed by a Friedman rank-sum test. The Boruta algorithm was used to select the taxa that have predictive power (R package "Boruta"). PICRUSt was used to predict the abundance of functional categories such as clusters of orthologous group (COG) categories and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways based on the 16S rDNA sequence data³³. All the statistical analyses were performed in R-3.0.2 (R Development Core Teams).

References

1. Sospedra, M. & Martin, R. Immunology of multiple sclerosis. *Annu Rev Immunol* **23**, 683–747 (2005).
2. Ascherio, A. Environmental factors in multiple sclerosis. *Expert Rev Neurother* **13**, 3–9 (2013).
3. Oksenberg, J. R. Decoding multiple sclerosis: an update on genomics and future directions. *Expert Rev Neurother* **13**, 11–19 (2013).
4. Fleming, J. & Fabry, Z. The hygiene hypothesis and multiple sclerosis. *Annals of neurology* **61**, 85–89 (2007).
5. Rook, G. A. Hygiene hypothesis and autoimmune diseases. *Clinical reviews in allergy & immunology* **42**, 5–15 (2012).
6. Nusrat, S., Gulick, E., Levinthal, D. & Bielefeldt, K. Anorectal dysfunction in multiple sclerosis: a systematic review. *ISRN neurology* **2012**, 376023 (2012).
7. Yacyshyn, B., Meddings, J., Sadowski, D. & Bowen-Yacyshyn, M. B. Multiple sclerosis patients have peripheral blood CD45RO+ B cells and increased intestinal permeability. *Digestive diseases and sciences* **41**, 2493–2498 (1996).
8. Gupta, G., Gelfand, J. M. & Lewis, J. D. Increased risk for demyelinating diseases in patients with inflammatory bowel disease. *Gastroenterology* **129**, 819–826 (2005).
9. Kimura, K. *et al.* Concurrence of inflammatory bowel disease and multiple sclerosis. *Mayo Clinic proceedings* **75**, 802–806 (2000).
10. Braniste, V. *et al.* The gut microbiota influences blood-brain barrier permeability in mice. *Sci Transl Med* **6**, 263ra158 (2014).
11. Jeraldo, P. *et al.* IM-TORNADO: A Tool for Comparison of 16S Reads from Paired-End Libraries. *PLoS One* **9**, e114804 (2014).
12. Miyake, S. *et al.* Dysbiosis in the Gut Microbiota of Patients with Multiple Sclerosis, with a Striking Depletion of Species Belonging to Clostridia XIVa and IV Clusters. *PLoS One* **10**, e0137429 (2015).
13. Cantarel, B. L. *et al.* Gut microbiota in multiple sclerosis: possible influence of immunomodulators. *J Investig Med* **63**, 729–734 (2015).
14. Walters, W. A., Xu, Z. & Knight, R. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett* **588**, 4223–4233 (2014).
15. Korkina, L., Kostyuk, V., De Luca, C. & Pastore, S. Plant phenylpropanoids as emerging anti-inflammatory agents. *Mini Rev Med Chem* **11**, 823–835 (2011).
16. Schogor, A. L. *et al.* Ruminant Prevotella spp. may play an important role in the conversion of plant lignans into human health beneficial antioxidants. *PLoS One* **9**, e87949 (2014).
17. Toh, H., Oshima, K., Suzuki, T., Hattori, M. & Morita, H. Complete Genome Sequence of the Equol-Producing Bacterium *Adlercreutzia equolifaciens* DSM 19450T. *Genome announcements* **1** (2013).
18. Moussa, L. *et al.* A low dose of fermented soy germ alleviates gut barrier injury, hyperalgesia and faecal protease activity in a rat model of inflammatory bowel disease. *PLoS One* **7**, e49547 (2012).
19. Confavreux, C., Hutchinson, M., Hours, M. M., Cortinovis-Tourniaire, P. & Moreau, T. Rate of pregnancy-related relapse in multiple sclerosis. Pregnancy in Multiple Sclerosis Group. *The New England journal of medicine* **339**, 285–291 (1998).
20. Moore, S. M. *et al.* Multiple functional therapeutic effects of the estrogen receptor beta agonist indazole-Cl in a mouse model of multiple sclerosis. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 18061–18066 (2014).
21. Spence, R. D. & Voskuhl, R. R. Neuroprotective effects of estrogens and androgens in CNS inflammation and neurodegeneration. *Frontiers in neuroendocrinology* **33**, 105–115 (2012).
22. Labbe, A., Ganopolsky, J. G., Martoni, C. J., Prakash, S. & Jones, M. L. Bacterial bile metabolising gene abundance in Crohn's, ulcerative colitis and type 2 diabetes metagenomes. *PLoS One* **9**, e115175 (2014).

23. Vavassori, P., Mencarelli, A., Renga, B., Distrutti, E. & Fiorucci, S. The bile acid receptor FXR is a modulator of intestinal innate immunity. *J Immunol* **183**, 6251–6261 (2009).
24. Gevers, D. *et al.* The treatment-naïve microbiome in new-onset Crohn's disease. *Cell host & microbe* **15**, 382–392 (2014).
25. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331**, 337–341 (2011).
26. Kang, D. W. *et al.* Reduced incidence of *Prevotella* and other fermenters in intestinal microflora of autistic children. *PLoS One* **8**, e68322 (2013).
27. Murri, M. *et al.* Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. *BMC medicine* **11**, 46 (2013).
28. Mukhopadhyay, I., Hansen, R., El-Omar, E. M. & Hold, G. L. IBD-what role do Proteobacteria play? *Nature reviews. Gastroenterology & hepatology* **9**, 219–230 (2012).
29. Jenq, R. R. *et al.* Regulation of intestinal inflammation by microbiota following allogeneic bone marrow transplantation. *J Exp Med* **209**, 903–911 (2012).
30. Bajaj, J. S. *et al.* Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. *American journal of physiology. Gastrointestinal and liver physiology* **303**, G675–G685 (2012).
31. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology* **12**, 87 (2014).
32. McDonald, W. I. *et al.* Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Annals of neurology* **50**, 121–127 (2001).
33. Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology* **31**, 814–821 (2013).

Acknowledgements

This study was sponsored by the Mayo Clinic Center for Individualized Medicine and the Mayo-Illinois Alliance for Technology Based Healthcare. We thank Dr. Kristin Inman (Mayo Clinic) for critically reading and editing the manuscript.

Author Contributions

Dr. J.C. performed the analysis and wrote the manuscript. Dr. N.C. conceptualized the study, performed analysis, and wrote the manuscript; Drs. K.R.K. and J.Z.Y. performed analysis; Dr. M.N., Ms. J.Z.Y. and Mr. D.H.L. helped with sample collection and performed experiments; Drs. M.M.P.S., M.N. and O.H.K. helped with collection of multiple sclerosis patients' meta data and interpretation, as well as revised the manuscript; Mr. J.C. helped with analysis; Drs. B.G.W., O.H.K., M.R., H.N. and J.M. helped with the study design and interpretation of the data; Drs. A.K.M., N.C. and J.M. conceptualized the study, designed the experiments, wrote the manuscript, and gave final approval of the manuscript to be published.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Chen, J. *et al.* Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. *Sci. Rep.* **6**, 28484; doi: 10.1038/srep28484 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Advanced MRI and staging of multiple sclerosis lesions

Martina Absinta^{1,2}, Pascal Sati¹ and Daniel S. Reich¹

Abstract | Over the past few decades, MRI-based visualization of demyelinated CNS lesions has become pivotal to the diagnosis and monitoring of multiple sclerosis (MS). In this Review, we outline current efforts to correlate imaging findings with the pathology of lesion development in MS, and the pitfalls that are being encountered in this research. Multimodal imaging at high and ultra-high magnetic field strengths is yielding biologically relevant insights into the pathophysiology of blood–brain barrier dynamics and both active and chronic inflammation, as well as mechanisms of lesion healing and remyelination. Here, we parallel the results in humans with advances in imaging of a primate model of MS — experimental autoimmune encephalomyelitis (EAE) in the common marmoset — in which demyelinated lesions resemble their human counterparts far more closely than do EAE lesions in the rodent. This approach holds promise for the identification of innovative biological markers, and for next-generation clinical trials that will focus more on tissue protection and repair.

More than a third of a century has elapsed since the publication of the first *in vivo* MRI study of multiple sclerosis (MS) lesions, a period during which the identification of such lesions has become our single most important diagnostic test for MS. Remarkably, the in-depth study of MS lesions by means of contemporary imaging techniques continues to uncover fundamental aspects of the pathophysiology of the disease¹. Multimodal imaging at high and ultra-high magnetic field strengths has yielded biologically relevant insights into lesion structure and blood–brain barrier (BBB) dynamics that are gradually narrowing the gap between the macroscopic view of the radiologist and the microscopic view of the pathologist.

In this Review, we explore how imaging findings can mirror the pathological development of the demyelinated MS lesion. We describe the patterns by which gadolinium leaks into the parenchyma, either centrifugally (outward from the central vein) or centripetally (inward from the lesion edge), highlighting the dynamism of the BBB in veins and capillaries as it responds to inflammatory stimuli. The introduction of ultra-high-field 7 T human MRI scanners, and of MRI techniques that are highly sensitive to paramagnetic substances, has enabled visualization of prominent central veins within most demyelinated MS lesions, and of paramagnetic rims at the lesion margin in some cases. We also review progress in the imaging of demyelination and remyelination.

Throughout this article, we parallel the results in humans with recent advances in 7 T imaging of a primate model of MS, experimental autoimmune

encephalomyelitis (EAE) in the common marmoset. The demyelinated brain and spinal cord lesions resemble the human counterpart far better in this model than in the rodent EAE model². Animal models offer the possibility to image at baseline and frequently after disease induction, thereby enabling investigation of early, subtle radiological changes within normal brain tissue before lesion formation³. Another important advantage of animal models is that the pathology underlying the MRI findings can be accessed once the experiment is ended. Lesions in the marmoset brain are often small, so high-resolution ($\leq 100 \mu\text{m}$ isotropic voxels) scans of the fixed brain, performed before processing of the tissue for histology, can be very useful (FIG. 1). Customized 3D-printed tissue holders and slicers facilitate matching of the histopathology with MRI, which greatly improves the interpretability of the data⁴. This technology, combined with serial MRI scanning before and after lesion development, can elucidate the cellular basis of development and repair of EAE lesions and, by extension, the histopathologically similar lesions in MS itself.

Taken together, the advances described in this article hold considerable promise for deepening our understanding of the biology of MS while opening up the possibility of next-generation proof-of-concept clinical trials mainly focused on tissue protection and repair.

Early development of MS lesions

Immunopathology. Breakdown of immunological tolerance to CNS myelin or myelin-like antigens is generally accepted to be the basis of MS pathogenesis. Routes of

¹Translational Neuroradiology Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, 10 Center Drive MSC 1400, Building 10 Room 5C103, Bethesda, Maryland 20852, USA.

²Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Via Olgettina 60, 20132, Milan, Italy.

Correspondence to D.S.R. reichds@ninds.nih.gov

doi:10.1038/nrneuro.2016.59
Published online 29 Apr 2016

Key points

- Recent research efforts have narrowed the gap between *in vivo* MRI and pathology with respect to multiple sclerosis (MS) lesion development and staging
- Demyelinated lesions in marmoset experimental autoimmune encephalomyelitis (EAE) resemble their human counterparts far better than do lesions in rodent EAE models
- Application of parallel MRI methods in MS and marmoset EAE facilitates interpretation of data collected in both conditions
- The perivenular topography of inflammatory demyelinating lesions might facilitate the diagnostic work-up for MS
- Smouldering lesions and meningeal inflammation are important features of chronic inflammation that should be targeted for *in vivo* imaging

physiological immune surveillance have been recognized to be involved early in the pathogenesis of both MS and EAE. Peripheral lymphocytes undergo trafficking, participate in immune surveillance and interact with CNS-resident antigen-presenting cells in the meninges (at the blood–leptomeningeal barrier), along ventricular ependymal surfaces (blood–choroid plexus barrier), within perivascular spaces (BBB), and in cervical lymph nodes⁵. Once they escape peripheral tolerance, autoreactive lymphocytes can be further activated and become hostile effectors (CD8⁺, CD4⁺ T-helper 1 (T_H1), CD4⁺ T_H17 and B lymphocytes)⁶, thereby precipitating demyelination and axonal injury, which are the typical features of MS.

In early disease, the most dramatic immunopathological events leading to focal areas of demyelination — the ‘lesion’ of MS — tend to occur at the level of the BBB in small or medium-size parenchymal veins, although early subpial cortical demyelination has also been recognized⁷. As they follow the venous vasculature, demyelinated MS lesions are commonly found around the ventricles (deep venous system, subependymal and medullary veins), at the grey–white matter junction (superficial venous system, especially cortical veins), within the optic nerve (branches of the central retinal vein), and in the brainstem (anteriorly and transversely oriented deep veins) and spinal cord (often posterior deep spinal veins). The spatial predilection and timing of a particular vein’s involvement remain fascinating research topics. In general, however, the process of rapid demyelination, which progressively spreads from and along a prominent parenchymal vein to involve the surrounding tissue, directly influences the macrostructure and shape (ovoid, ellipsoidal, triangular or U-shaped) of MS lesions (FIG. 2).

Human imaging. As demyelinated lesions in the brain and spinal cord are easy to detect on long echo-time MRI pulse sequences, such as T2-weighted and T2-fluid-attenuated inversion recovery (FLAIR) sequences, these techniques have been implemented in everyday clinical practice and are integral to current MS diagnostic criteria^{8,9}. Though recognized in pathological studies as early as the nineteenth century¹⁰, the perivenular topography of MS lesion development has only recently been directly visualized *in vivo*, using high-resolution susceptibility-sensitive imaging at a variety of magnetic

field strengths (see [Supplementary information S1 \(video\)](#))^{11–18}. The reasons for the characteristic prominence of the central vein within MS demyelinated lesions in comparison to surrounding brain parenchyma are still open to debate, and include an elevated concentration of deoxyhaemoglobin (higher oxygen extraction at the site of inflammation) and changes in vessel diameter (slower venous flow, postinflammatory scarring processes of the vein wall)^{19–21}.

Far from being a mere morphological feature of MS lesions, the ‘central vein sign’ has potential relevance for the diagnostic work-up^{9,22}, as it is consistently described as atypical in small vessel disease^{15,16,23,24} and migraine²⁵, and is rarely observed in other immunological conditions associated with brain lesions (for example, vasculitis and neuromyelitis optica spectrum disorder)²⁶. Studies conducted in large populations will be required to ascertain the true positive predictive value of the central vein sign.

Marmoset imaging. Various EAE immunization protocols have been described in marmosets^{27–30}, but in this Review we will focus primarily on disease induced by subcutaneous injection of human white matter homogenate with an adjuvant^{3,27}, which in our hands produces a disease that bears strong similarities to relapsing–remitting MS. As in MS, MRI can detect demyelinated lesions that form in the CNS of marmosets with EAE^{31–33}. EAE lesions in the marmoset brain present with similar features to MS lesions, including distribution throughout the CNS with a preferred location in the subcortical white matter^{3,34,35}, uptake of gadolinium contrast agent^{3,36}, and perivenular topography³⁷ (FIG. 1). Interestingly, the veins around which most of the lesions form are already visible on scans before disease induction, supporting the notion that the inflammation associated with EAE affects previously normal veins. These superficial medullary veins drain the white matter lesions and pass through the cortex into the subarachnoid space. The close relationship between these veins and cerebrospinal fluid within their perivascular spaces is likely to have an important role in lesion development.

Before abrupt opening of the BBB

Immunopathology. A prevalent adaptive immune response to myelin-like antigens triggers an aggressive, scavenger-type response of the innate immune system against CNS myelin. This response is characterized by a classic triad: first, dramatic opening of the BBB; second, blood-derived inflammatory cell (monocyte–macrophage) invasion; and third, macrophage-mediated stripping of myelin sheaths from axons^{38,39}. The temporal sequence of the earliest pathological events has not been completely unravelled, and remains a subject of active investigation in animal studies. Perivascular inflammatory cuffing is the most prominent pathological feature associated with the newly forming lesion, and is characterized by aggregates of lymphocytes and monocytes confined within the perivascular space of the postcapillary venule and not yet breaching the glia limitans of the BBB (parenchymal basement membrane and

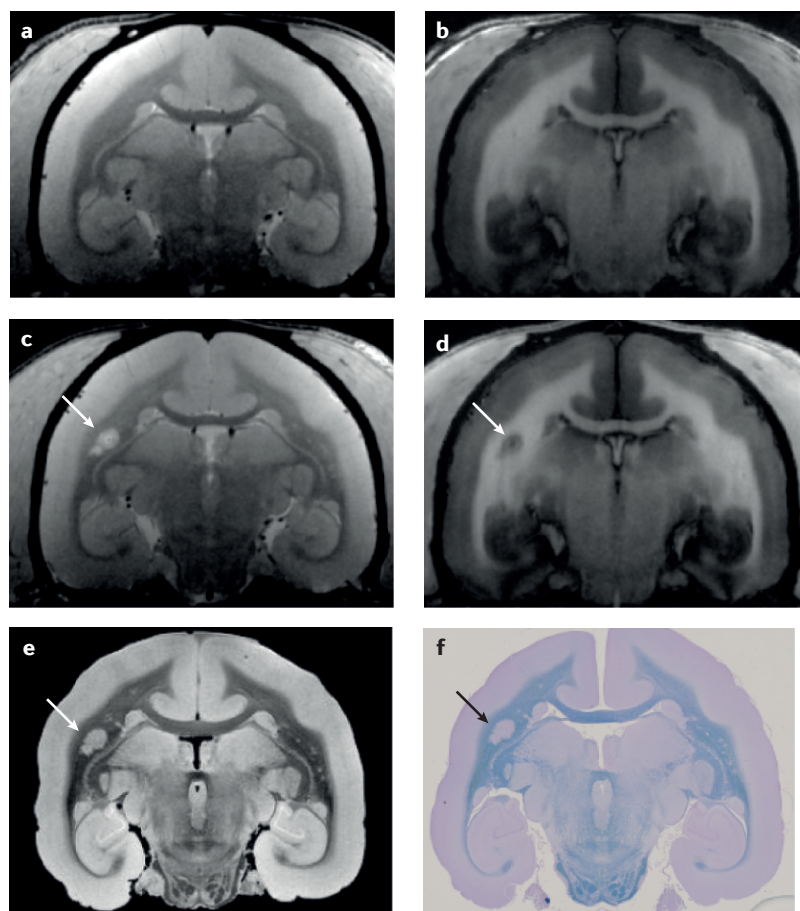


Figure 1 | MRI and histology of the marmoset brain. A common marmoset housed at the NIH Intramural Research Program facility was studied in accordance with the standards of the American Association for Accreditation of Laboratory Animal Care and the National Institute of Neurological Disorders and Stroke's Animal Care and Use Committee. **a,b** | Baseline 7 T MRI showing a healthy brain on proton density-weighted (**a**) and T1-weighted (**b**) images. After MRI, the animal was immunized with 200 µg of fresh-frozen human white matter homogenate, and started developing white matter lesions soon after. **c,d** | These lesions are visible on follow-up proton density-weighted (**c**) and T1 weighted (**d**) MRI scans performed before termination of the experiment. Note that a central vein within the lesion (arrow) can be discerned on the proton density image (**c**). **e,f** | 7 T postmortem 100 µm isotropic MRI after necropsy (**e**) was used to guide tissue processing for histology (**f**), allowing precise correlation between MRI and histopathology (arrows). Complete demyelination, as seen on Luxol fast blue staining (**f**), matches the radiological appearance of the lesion both *in vivo* and post mortem (**c, d, e**).

astrocytic end-feet). Other notable events include variable immunoglobulin and complement deposition (the distinguishing feature of immunopattern II compared with immunopattern I)⁴⁰; parenchymal microglial and astrocytic activation; perivascular swollen or apoptotic oligodendrocytes with still-intact myelin (the so-called prephagocytic or (p)reactive lesion)^{41,42}; and early axonal and synaptic dysfunction^{38,39,43,44}.

Human imaging. The early pathological findings in both MS and EAE are consistent with the *in vivo* detection of subtle radiological changes in the normal-appearing white matter days to weeks before the parenchymal gadolinium enhancement that defines the radiological onset of the impending MS lesion^{45–52}. These data,

derived mostly from quantitative or semiquantitative MRI studies that are pathologically nonspecific and only statistically significant at the group level, support the concept of a short-term temporal dissociation between the earliest immunological events in lesion development and the dramatic opening of the BBB (captured on MRI as parenchymal leakage of gadolinium) and subsequent demyelination. Recently, taking advantage of high-resolution imaging and frequent scanning, we have directly visualized focal alterations of MRI signal, on T2-FLAIR and T2*-weighted images, around the central vein of the future MS lesion¹⁸. Specifically, in some cases, before the overt onset of the radiologically defined demyelinated lesion, the ‘inflamed’ central vein enhances in isolation, perhaps owing to higher BBB permeability with entrapment of the contrast agent within the perivascular space¹⁸.

Marmoset imaging. In a study published in 2014 (REF. 3), we monitored the status of BBB permeability by mapping the longitudinal relaxation time (‘T1’) before and after injection of a triple dose of gadolinium-based MRI contrast agent. We performed weekly MRI scans, including T1 mapping, as well as proton density-weighted and T2-weighted imaging, after induction of EAE in marmosets. We showed that the BBB becomes focally and increasingly permeable starting approximately 4 weeks before lesion appearance. We also demonstrated that areas of abnormal BBB permeability observed at the time of sacrifice corresponded to inflammatory nodules without associated demyelination. Those nodules consisted of small clusters of perivascular and parenchymal, Iba1-positive, activated microglia and macrophages in association with variable numbers of perivascular lymphocytes. Our results suggest that early changes in vascular permeability, associated with perivascular inflammatory cuffing and parenchymal microglial activation, might precede the formation of new lesions.

BBB dynamics in early MS lesions

Immunopathology. The BBB is recognized to be a highly organized structure, composed of endothelial cells, pericytes and astrocytic end-feet, all of which partially isolate the peripheral blood circulation from the CNS parenchyma⁵³. The endothelial and parenchymal basement membranes, consisting of collagen IV, fibronectin and laminin, enclose the ‘perivascular’ space, in which — as stated above — normal and pathological immune surveillance takes place. In early MS lesions, autoreactive and activated lymphocytes directly contribute to impairment of the BBB of the postcapillary venules, characterized initially by moderately increased permeability and eventually by dramatic structural disruption^{38,39,44,54}. Activated CD4⁺ lymphocytes can secrete the cytokines IFN-γ and tumour necrosis factor (TNF), which upregulate endothelial adhesion molecules (VCAM1 and α1 integrin) and chemokines, thereby inducing massive recruitment of monocytes and allowing transendothelial migration of these cells into the perivascular space and, subsequently, the parenchyma. Structural disruption of the BBB is mediated by the secretion of matrix

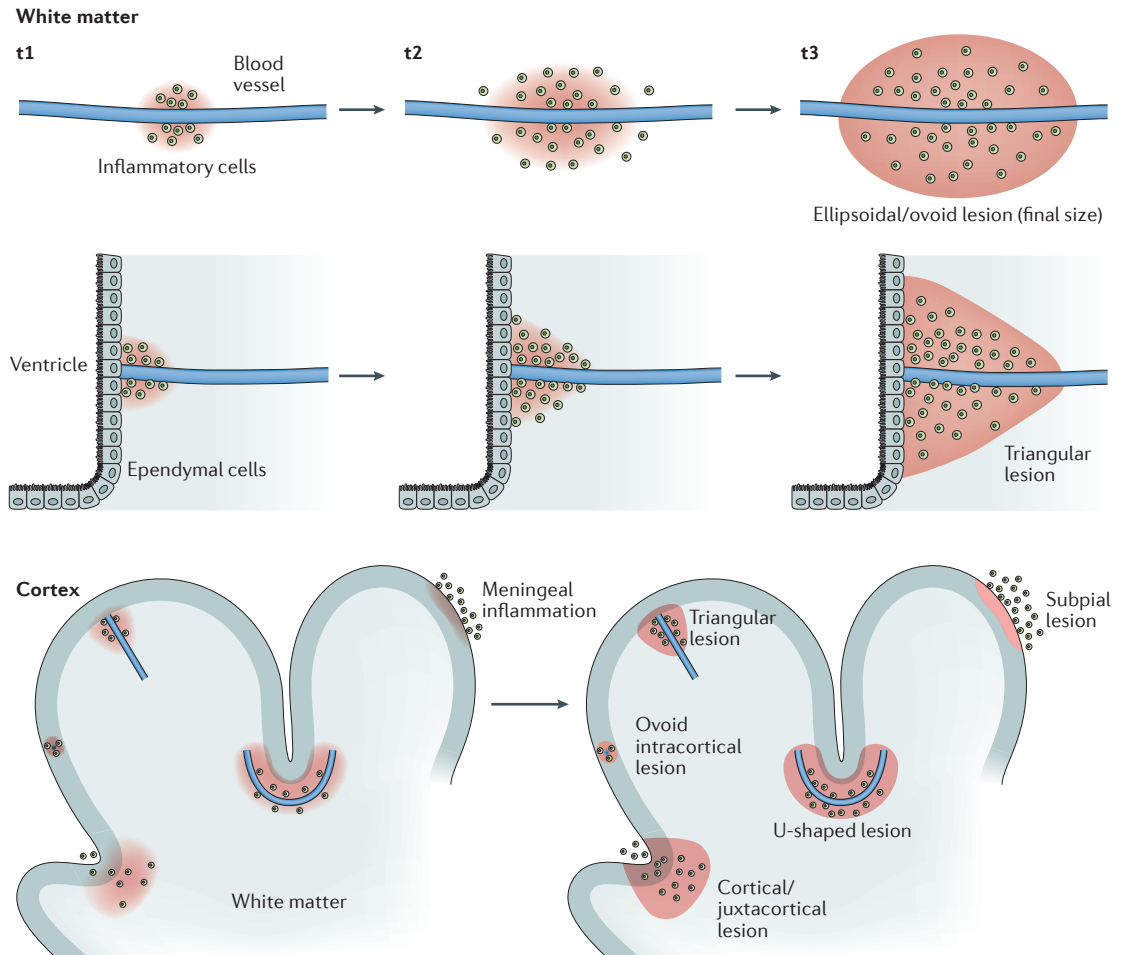


Figure 2 | Perivenular topography of demyelinated MS lesions in white and grey matter. In the newly forming MS lesion, the distribution of myelin-scavenging inflammatory cells, spreading along and outward from an inflamed central vein, largely dictates the final lesion configuration. An ellipsoidal and/or rounded configuration is typical of MS lesions in the white matter. Of note, the shape of MS lesions can be also affected by the macrostructure of the surrounding tissue. For example, both the ependymal wall of the ventricles and the cortex may act as barriers to the spread of myelin-scavenging inflammatory cells. Depending on the trajectory of the vein and its position relative to these structures, lesions might assume a triangular shape, for periventricular and leukocortical–juxtacortical lesions, or a U shape, for leukocortical–juxtacortical lesions. The lower panel shows the various types of cortical demyelination. In addition to vasocentric demyelination, which occurs in purely intracortical and leukocortical lesions, plaque-like demyelination can affect the subpial layers of the cortex and is thought to be directly triggered by leptomeningeal inflammation. MS, multiple sclerosis; t, time point.

metalloproteinases that can damage endothelial tight junctions and basement membranes, causing leakage of serum proteins and water (vasogenic oedema) into the white matter^{55,56}. Impaired breakdown of leaking fibrinogen, and its breakdown product fibrin, might also contribute to the inflammatory demyelinating cascade^{57,58}. This highly inflammatory environment eventually leads to disorganization of the BBB's outer shell (the astrocytic end-feet that comprise the glia limitans).

Human imaging. The striking radiological appearance of newly forming MS lesions is commonly accepted to be related to leakage into the parenchyma (and, therefore, visualization on T1-weighted images) of intravenously injected contrast agents, which are usually gadolinium-based^{59,60}. This leakage reflects the short-term structural

opening of the BBB — specifically, damage to tight junctions and basement membranes — that occurs at lesion onset.

Dynamic contrast-enhanced MRI (DCE-MRI) at high magnetic field strength is a technique in which T1-weighted images are quickly acquired before, during and at various time points after gadolinium injection. Through use of this technique, the spatiotemporal dynamics of BBB opening in active MS lesions can be dissected *in vivo*, thereby providing complementary insights to the immunopathological studies. In this dynamic view, in essentially all early active demyelinating lesions, gadolinium leaks initially from the central vein and flows outward, filling the lesion within a few minutes. On static images acquired 5–10 min after contrast injection — the typical situation in clinical practice

— such lesions would be characterized as having ‘nodular’ enhancement. From a dynamic point of view, this process has been termed centrifugal (outward) contrast enhancement^{61,62}.

After a few days to a few weeks, many of the larger lesions exhibit a shift in the pattern of gadolinium enhancement. In such cases, the expanding inflammatory edge enhances first, and gadolinium then flows centripetally (inward) toward the demyelinated lesion centre^{17,61,62} (see [Supplementary information S2, S3](#) (videos)). When this shift occurs, the appearance of the lesion on static images might be either a nodule or a ring, depending on several factors. Especially important among these factors is the time elapsed between contrast injection and MRI acquisition, but the location of the lesion is also important in determining whether the ring is complete or open: an open ring is typically seen when the lesion abuts grey matter or the ventricles. Centrifugal enhancement clearly supports the centrality of perivenular pathogenic events at lesion onset. The significance of centripetal enhancement is less certain, and we hypothesize that this feature might be part of the physiological — though detrimental if unrestrained — wound-healing inflammatory reaction of the CNS parenchyma to focal demyelination.

Marmoset imaging. The dramatic opening of the BBB can also be visualized in active EAE lesions, using T1-weighted images obtained after intravenous injection of gadolinium-based contrast agents. Like MS lesions, active EAE lesions usually display nodular enhancement on static postcontrast T1-weighted images. To our knowledge, open-ring enhancement has not been reported in marmoset EAE. The spatiotemporal dynamics of the gadolinium leakage in active EAE lesions have not yet been fully characterized, mainly due to the limitations imposed by the smaller dimensions of EAE lesions and the higher cardiac output of marmosets compared with humans. Nonetheless, a recent quantitative MRI study has provided some insights into the dynamics of BBB leakage over the span of a few weeks³. Following a gradual increase in permeability that occurs early in lesion development, the BBB reaches its maximum permeability once the lesion becomes visible on all types of images (T1-weighted, T2-weighted, T2*-weighted and proton density-weighted, as well as T2-FLAIR). After the lesion appears, the permeability of the BBB gradually decreases back to near-normal levels, signifying a restored BBB, in approximately 2–4 weeks³.

Macrophage-mediated demyelination

Immunopathology. Stripping of myelin from axons and phagocytosis of myelin debris are mediated by blood-derived macrophages and activated CNS-resident microglia, which together constitute the prevalent cellular population within the active demyelinating lesion^{38,63–65}. These cell types are difficult to distinguish even at the tissue level (although progress has recently been made on this front⁶⁶), and here we will use the term ‘macrophage’ for simplicity.

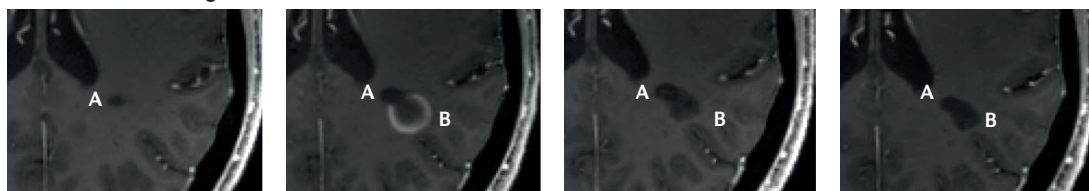
The proinflammatory status of macrophages probably contributes to further tissue damage via free radical formation and secretion of proteases and cytokines, including IFN- γ and TNF. Macrophages track the spatiotemporal pattern of BBB opening and demyelination described above. Thus, they initially distribute throughout the lesion, starting from the centre, and they subsequently become more concentrated at the lesion margin (so-called ‘late active’ lesions). Interestingly, the efficiency by which macrophages clear myelin debris seems to decrease with age⁶⁷. In addition, in an attempt to limit amplification of oxidative stress and mitochondrial energy failure during early lesion development, an excess of free iron, probably released from oligodendrocytes during the process of demyelination and/or mediated by endothelial transferrin receptor upregulation, is buffered, mainly by macrophages and astrocytes^{68–70}. Persistently demyelinated axons are more prone to degenerate in an inflammatory environment^{71,72}, which might explain the varying degrees of axonal injury and loss that coexist with demyelination in MS lesions.

Human imaging. A persistent issue in imaging of MS lesions *in vivo* is related to the fact that inflammation, demyelination and axonal loss within lesions are not clearly distinguishable on T2-weighted and T1-weighted images. This lack of specificity has prompted a number of attempts to develop MRI approaches that are more pathologically interpretable, such as quantification of the short T2 component in multi-echo T2 relaxation experiments, which seems to correlate with water selectively enclosed between myelin bilayers (so-called ‘myelin water imaging’)^{73–76}.

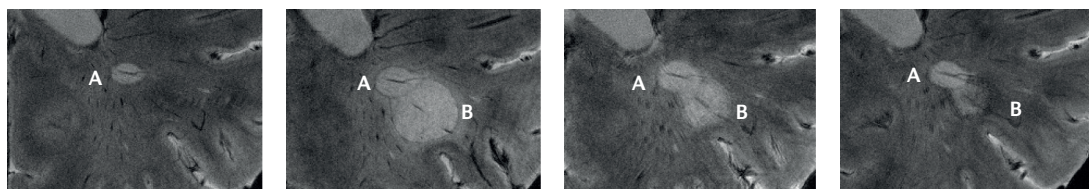
Complementary work on the spatiotemporal dynamics of the BBB has shown that the front of demyelination and inflammation can be captured by coupling DCE-MRI with susceptibility-based MRI at 7 T (REF. 17). These experiments have shown that in centripetally enhancing lesions, the initial opening of the BBB at the lesion edge clearly colocalizes with a paramagnetic rim on phase images — a novel endogenous marker of inflammation¹⁷ (see [Supplementary information S2](#) (video) and [FIG. 3](#)). Other investigations showed that on 3 T phase images, visualization of phase MRI changes was better overall in young lesions in early disease than in older lesions^{51,77–79}. Pathological features of the younger lesions that might explain this observation include more-prominent demyelination in comparison to axonal loss, and high inflammatory activity with attendant paramagnetic substances, such as free radicals and iron.

Marmoset imaging. By combining serial conventional imaging and histopathology, we have demonstrated that <1-week-old EAE lesions surround a central inflamed blood vessel that is cuffed with T cells and macrophages³. Perivascular B cells and blood-derived MRP14⁺ early activated macrophages are also observed, albeit less frequently than T cells and MRP14⁺ macrophages, and macrophage–microglia-associated demyelination and disrupted axons are seen throughout the lesion. Lesions 1–5 weeks old continue to enlarge, and

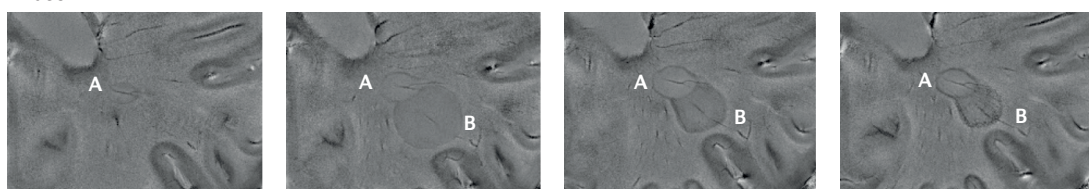
Postcontrast T1-weighted



T2*-weighted magnitude



Phase



0 6 7 12
Time (months)

Figure 3 | MS lesions with paramagnetic rims. An untreated 49-year-old woman with progressive MS and radiological relapses underwent serial MRI evaluations at 7 T under an institutional review board-approved natural history protocol at the NIH. Lesion A is a pre-existing, nonenhancing lesion with a paramagnetic rim visible on both T2*-weighted magnitude and phase images. Lesion B, a new centripetally enhancing MS lesion (open ring on postcontrast T1-weighted images; Supplementary information S2 (video)), appears adjacent to lesion A at month 6. Both lesions exhibit a central vein on T2*-weighted magnitude and phase images. In lesion B, the peripheral T2*-weighted magnitude and phase rim persists after resolution of contrast enhancement (1 and 6 months after lesion onset). Of note, lesion B shrinks over time after enhancement resolves, whereas lesion A nearly triples in size, even in the absence of visible contrast enhancement on postcontrast T1-weighted images. By month 12, the persistent T2*-weighted magnitude and phase rims in both lesions have become darker and thicker, suggesting accrual of paramagnetic substances at the lesion edge. Pathologically, the rim possibly results from macrophages loaded with iron, and lesions with this radiological feature might, therefore, correspond to the so-called 'chronic active' or 'smouldering' lesions. MS, multiple sclerosis.

active demyelination is observed at the lesion edges. Interestingly, early activated macrophages are only found at the leading edge of demyelination³.

Conventional MRI lacks specificity toward tissue pathology in EAE lesions, as in MS lesions, and inflammation, demyelination and axonal loss cannot be differentiated *in vivo*³¹. In one study, quantitative MRI findings correlated with macrophage infiltration, but demyelination was a potential confounder³³. Recently, an advanced MRI technique, based on complex analysis of the T2* decay curve, was used for *in vivo* detection of the myelin water content in white matter of healthy marmosets⁷⁵. However, the application of this technique to track demyelination and remyelination remains challenging, as T2* decay curves can be strongly affected by oedema in the extracellular space, iron accumulation in macrophages, and the orientation of white matter tracts relative to the magnetic field of the MRI system. To date, a paramagnetic rim on phase images has not been observed in marmoset EAE lesions, and whether this is due to lack of sensitivity of the applied MRI techniques or to biological differences remains uncertain.

Lesion evolution after BBB restoration

Immunopathology. The processes of BBB closure and resolution of inflammation are recognized as potential sources of new therapeutic targets in MS, but the underlying mechanisms have not been completely unravelled. An emerging idea is that inflammation in MS, while clearly driving demyelination and tissue destruction, can also trigger healing, albeit dysregulated and, hence, either incomplete or ineffective. In this context, an analogy has been drawn between extra-CNS wound healing and the process of repair within MS lesions⁸⁰. The initiation of healing seems to reside in the extreme plasticity of macrophages, the functional phenotypes of which run the gamut from destructive ('classically activated') to restorative ('alternatively activated') in the face of environmental stimuli that may vary over time. In the realm of remyelination^{81–85}, recent data indicate that in early lesions, classically activated macrophages (sometimes called 'proinflammatory' macrophages) can induce the recruitment of oligodendrocyte precursors, whereas alternative ('anti-inflammatory') activation is required for differentiation of these precursor cells

into myelinating oligodendrocytes⁶⁷. Remyelination remains the subject of extensive investigations that are outside the scope of this Review^{84–89}.

Human imaging. After the resolution of gadolinium enhancement, usually 2–8 weeks following its emergence, MS lesions are radiologically defined as ‘chronic’. The lesions remain clearly visible on both T2-weighted and T1-weighted images, and the extent of T1 hypointensity seems to correlate with axonal degeneration in the white matter^{90,91}. Unfortunately, as discussed above, clinical MRI sequences are unable to directly image remyelination. After oedema resolves, typically 1–3 months after lesion onset, tracking of longitudinal signal intensity on T1-weighted, T2-weighted, proton density-weighted and magnetization transfer-weighted sequences, as well as through lesion volume measurement, remains the established strategy to assess short-term lesion repair^{92,93}.

Marmoset imaging. The subacute EAE lesions described above show a decrease in BBB permeability in their core area over approximately 4 weeks³. By 5 weeks, the lesions display markedly reduced BBB permeability and inflammation, and the macrophages disappear. These lesions are still visible on MRI but become smaller over time — a finding that is consistent not only with a reduction in inflammation but also with potential repair, including remyelination. As most marmoset experiments end within weeks to months of lesion development, the natural history and pathology of chronic lesions has not been extensively studied, and is an enticing topic for future research.

Smouldering inflammation

Immunopathology. On histological sections, chronic lesions appear hypocellular and demyelinated, with a varying degree of axonal loss. However, an inflammatory infiltrate, wherein macrophages and activated microglia contain early or late myelin degradation products, characterizes the margins of a variable percentage (20–40%) of demyelinated lesions (variously termed ‘chronic active’, ‘slowly expanding’, or ‘smouldering’)^{63,64,68,94,95}. On the basis of autopsy studies, this lesion category is especially common in people at the transition to clinically progressive disease and those with long disease duration, supporting the notion that these lesions contribute to long-term and clinically important tissue damage^{68,94,95}.

Human imaging. Chronic inflammation is trapped behind an intact BBB and, hence, remains invisible to current gadolinium-based MRI. Through the use of high-resolution, noncontrast, susceptibility-based imaging, a paramagnetic rim has recently been detected at the margin of some nonenhancing MS lesions. Initial MRI–pathology investigations suggested that this radiological finding might aid identification of chronic active/smouldering lesions *in vivo*. Steady accumulation of macrophages at the lesion edge, in association with paramagnetic inflammatory species related to free radical production and/or intracellular iron accumulation, might explain the paramagnetic signal observed *in vivo*^{12,17,78,96–100} (FIG. 3).

In the context of early lesion evolution, the subset of centripetally enhancing lesions with a phase rim that persists after the closure of the BBB tend to exhibit poor lesion recovery in terms of lesion volume and T1 hypointensity at year 1, suggesting impairment of early repair and/or permanent tissue injury¹⁰¹ (FIG. 3). If confirmed, this feature will constitute a valuable imaging biomarker, not only to better stage MS lesions *in vivo*, but also to uncover specific components of MS-related hidden inflammation. Similarly, peripheral macrophages experimentally labelled with ultrasmall superparamagnetic iron oxide particles (USPIO) have been tracked at the edges of demyelinated lesions with an apparently intact BBB (that is, lesions with no observable leakage of traditional gadolinium-based contrast agents)^{102–104}. A direct comparison between USPIO-based and susceptibility-based imaging is currently lacking.

Marmoset imaging. Bert ‘t Hart and colleagues applied conventional imaging (T2-weighted and T1-weighted with gadolinium injection) to a group of marmosets reaching the chronic stage of EAE disease³¹. Among the chronic lesions, late active demyelinating lesions (with macrophages containing myelin proteolipid protein degradation products located at the edge) were observed together with inactive demyelinated and remyelinated lesions, as confirmed using histopathology³¹. However, the MRI scans performed in this study could not differentiate these chronic active lesions from their inactive counterparts. USPIO-based MRI, which has been successful in detecting infiltrated macrophages in rodent EAE^{105–107}, could prove useful for identifying chronic active lesions in marmoset EAE. Another approach would be to perform serial susceptibility-based imaging to detect phase rims in those chronic lesions. These strategies represent exciting targets for future research.

Caveat

Four different initial immunological patterns are associated with acute myelin damage and BBB opening, supporting the idea that MS is a collection of different pathophysiological entities^{40,108,109}. What has been described in this Review — the staging of white matter MS lesions and its MRI correlates — probably refers to immunological patterns I and II (myelin-autoimmune hypothesis), which together are the most common types of active MS lesions⁴⁰. No imaging correlate has yet been shown to definitively identify lesions with immunological pattern III, in which the prime mover is thought to be toxic and/or hypoxic injury to mature oligodendrocytes⁴⁰.

Cortical and leptomeningeal involvement

Immunopathology. Cortical and leptomeningeal involvement has recently been recognized as a key feature in MS pathogenesis (early disease with clinical relapses and remission)^{7,110} and related neurodegeneration (thought to be largely responsible for clinical progression)¹¹¹. Like the white matter, the cortex can display demyelination, but with much less pronounced inflammation and, importantly, without

abrupt opening of the BBB^{7,112–114}. Remyelination may be enhanced within the cortex in comparison with the white matter^{84,115}.

Topographically, cortical lesions develop along the pial surface ('subpial lesions') or around cortical venules that drain into the subarachnoid space (small intracortical and leukocortical demyelinated lesions)^{112,116} (FIG. 2). This topography supports the hypothesis that leptomeningeal inflammation, involving effector lymphocytes, cytotoxic diffusible mediators and immunoglobulins, is strictly related to cortical lesion development via direct damage of the subpial layers of the cortex and/or inflammation spreading into the parenchyma⁷. Subpial cortical demyelination has been characterized as specific to MS-related pathological processes¹¹⁷. Of note, aggregates of lymphocytes resembling lymphoid follicles have been recognized in the leptomeninges in patients with progressive MS, and have been judged to reflect an extreme form of compartmentalized, long-lasting inflammation^{7,118–123}. The interplay among leptomeningeal inflammation, cortical demyelination and neurodegeneration, as well as more generally between tissue damage in white and grey matter, is still under investigation^{111,124,125}.

Human imaging. Cortical lesions have been reported *in vivo* at all stages of MS (though more frequently in progressive MS), but subsequent autopsy findings indicate that their numbers are profoundly underestimated on MRI^{126–128}. Of note, cortical lesions do not enhance after gadolinium administration, which at present precludes their staging *in vivo*. Double inversion recovery¹²⁶, phase-sensitive inversion recovery¹²⁹, and

high-resolution T2*-weighted^{130–132} and magnetization-transfer-weighted sequences^{133,134} can detect some of these lesions, especially those with transcortical involvement and/or demyelination of the underlying white matter ('leukocortical lesions'). Subpial demyelination (affecting only the superficial cortical layers) and small intracortical lesions have been visualized more reliably on high-resolution postmortem images (FIG. 4), but are only sporadically reported *in vivo*^{130–132}. Cortical demyelination shows positive correlations with disability and cognitive impairment scores, thereby underlining the clinical importance of improved detection of this feature^{131,135–139}. Together with periventricular, infratentorial, spinal cord and optic nerve lesions, cortical/juxtacortical lesions contribute to the recently proposed revised 'dissemination in space' MRI criteria for the diagnosis of MS⁹.

With respect to leptomeningeal inflammation, a recent study, in which *in vivo* and postmortem MRI findings were correlated with pathology, reported the prevalence of subarachnoid-space abnormalities in a cohort of ~300 people with MS on both 3 T and 7 T postcontrast T2-FLAIR images¹⁴⁰. Spatially confined disruption of the blood–leptomeningeal barrier, manifesting as foci of gadolinium enhancement, was identified in ~35% of individuals with progressive MS and ~25% of all patients studied (FIG. 4). The presence of leptomeningeal enhancement correlated with whole-brain and cortical atrophy¹⁴⁰. In autopsy cases, these foci colocalized with perivascular inflammatory infiltrates — specifically, lymphocytes and macrophages clustered around leptomeningeal vessels and along intracortical perivascular spaces — and adjacent subpial cortical demyelination¹⁴⁰.

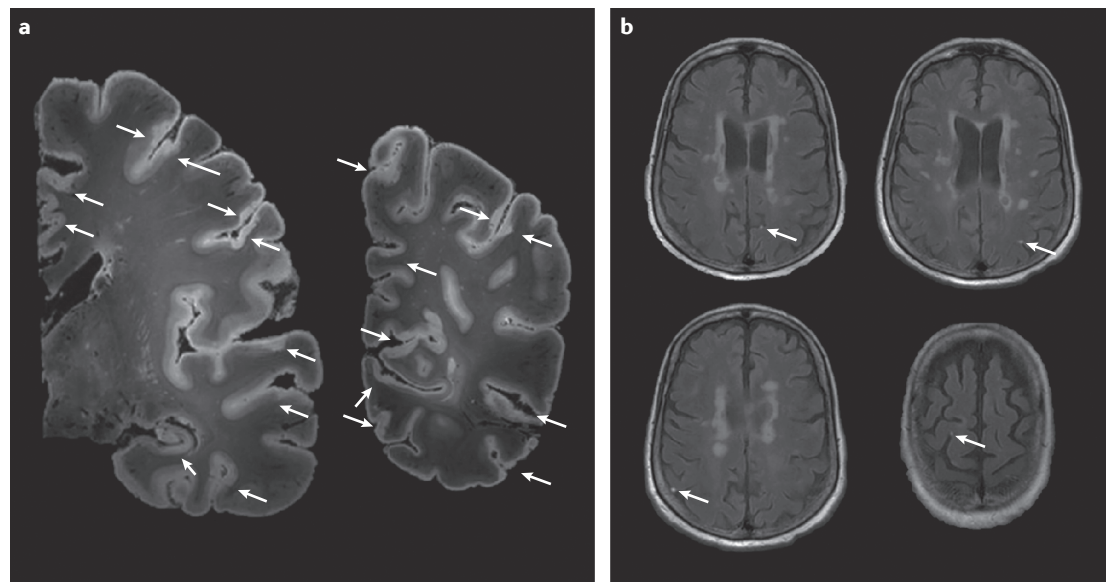


Figure 4 | Cortical and leptomeningeal involvement. **a** | 7 T postmortem MRI (two representative coronal slices of the left hemisphere, T2*-weighted gradient-echo sequence, 64 nl isotropic voxels) in a 66-year-old woman with progressive multiple sclerosis and disease duration (time from symptom onset to death) of 20 years. The arrows indicate widespread involvement of the cortex in the demyelination process, including subpial cortical and leukocortical lesions. Cortical lesions often face each other across sulci, consistent with the idea that leptomeningeal inflammation has a direct pathogenic role in cortical lesion development. **b** | Several foci of leptomeningeal contrast enhancement (arrows) were apparent in this patient on postcontrast 3 T MRI *in vivo*.

Marmoset imaging. Demyelinated cortical lesions similar to those in MS (leukocortical, intracortical and subpial) are known to affect the marmoset EAE brain, as shown in several histopathological studies^{141–144}. Clusters of T lymphocytes and plasma cells have recently been observed in the meninges overlying subpial cortical demyelination¹⁴⁵, suggesting that local meningeal lymphocytic and plasma cell infiltration also contributes to the pathogenesis of subpial demyelination in marmoset EAE. However, *in vivo* MRI visualization of cortical demyelination and meningeal inflammation has not yet been reported in this animal model.

Conclusions

In this Review, we have summarized recent research efforts to narrow the gap between pathology and imaging in MS lesion development and staging. We have also presented parallel efforts in marmoset EAE, an animal model bearing strong immunological, pathological and radiological similarities to MS. In marmoset EAE, the pathological basis of developing and repairing lesions can be investigated using a combination of serial *in vivo* MRI imaging and histology, and the results of these investigations have strong relevance for

our understanding of similar processes in MS. Some observations from human imaging, such as open-ring enhancement, phase rims, meningeal enhancement and cortical lesions, have not been consistently reported in marmoset EAE, and certain observations from marmoset imaging, such as consistently increased permeability of the BBB in the weeks preceding focal demyelination, have not been described in humans. These differences may be attributable in large part to technical factors, but biology could also contribute.

The parallels reviewed here suggest that the combination of human MS and marmoset EAE research is a promising pathway to achieve a deeper understanding of the pathobiology of BBB dynamics, chronic inflammation, and mechanisms of repair and remyelination, together with the development of new MRI techniques to detect and monitor their evolution *in vivo*. The ability to image both marmosets and humans at the same field strength (7 T) enables ready investigation and translation of new MRI markers from bench to bedside, and vice versa. Finally, this unique approach holds promise for the identification of novel biological targets and next-generation clinical trials that will focus on tissue protection and repair — a major unmet need in MS.

- Filippi, M. *et al.* Association between pathological and MRI findings in multiple sclerosis. *Lancet Neurol.* **11**, 349–360 (2012).
- t Hart, B. A., van Kooyk, Y., Geurts, J. J. & Gran, B. The primate autoimmune encephalomyelitis model; a bridge between mouse and man. *Ann. Clin. Transl. Neurol.* **2**, 581–593 (2015).
- Maggi, P. *et al.* The formation of inflammatory demyelinated lesions in cerebral white matter. *Ann. Neurol.* **76**, 594–608 (2014).
- Guy, J. R. *et al.* Custom fit 3D-printed brain holders for comparison of histology with MRI in marmosets. *J. Neurosci. Methods* **257**, 55–63 (2016).
- Ransohoff, R. M. & Engelhardt, B. The anatomical and cellular basis of immune surveillance in the central nervous system. *Nat. Rev. Immunol.* **12**, 623–635 (2012).
- Bartholomäus, I. *et al.* Effector T cell interactions with meningeal vascular structures in nascent autoimmune CNS lesions. *Nature* **462**, 94–98 (2009).
- Lucchinetti, C. F. *et al.* Inflammatory cortical demyelination in early multiple sclerosis. *N. Engl. J. Med.* **365**, 2188–2197 (2011).
- Polman, C. H. *et al.* Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann. Neurol.* **69**, 292–302 (2011).
- Filippi, M. *et al.* MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. *Lancet Neurol.* **15**, 292–303 (2016).
- Rae-Grant, A. D., Wong, C., Bernatowicz, R. & Fox, R. J. Observations on the brain vasculature in multiple sclerosis: a historical perspective. *Mult. Scler. Relat. Disord.* **3**, 156–162 (2014).
- Tan, I. L. *et al.* MR venography of multiple sclerosis. *AJNR Am. J. Neuroradiol.* **21**, 1039–1042 (2000).
- Hammond, K. E. *et al.* Quantitative *in vivo* magnetic resonance imaging of multiple sclerosis at 7 Tesla with sensitivity to iron. *Ann. Neurol.* **64**, 707–713 (2008).
- Tallantyre, E. C. *et al.* Demonstrating the perivascular distribution of MS lesions *in vivo* with 7-Tesla MRI. *Neurology* **70**, 2076–2078 (2008).
- Sati, P., George, I. C., Shea, C. D., Gaitan, M. I. & Reich, D. S. FLAIR*: a combined MR contrast technique for visualizing white matter lesions and parenchymal veins. *Radiology* **265**, 926–932 (2012).
- Tallantyre, E. C. *et al.* Ultra-high-field imaging distinguishes MS lesions from asymptomatic white matter lesions. *Neurology* **76**, 534–539 (2011).
- Kau, T. *et al.* The 'central vein sign': is there a place for susceptibility weighted imaging in possible multiple sclerosis? *Eur. Radiol.* **23**, 1956–1962 (2013).
- Absinta, M. *et al.* Seven-tesla phase imaging of acute multiple sclerosis lesions: a new window into the inflammatory process. *Ann. Neurol.* **74**, 669–678 (2013).
- Absinta, M. *et al.* Direct MRI detection of impending plaque development in multiple sclerosis. *Neurol. Neuroimmunol. Neuroinflamm.* **2**, e145 (2015).
- Gaitan, M. I., de Alwis, M. P., Sati, P., Nair, G. & Reich, D. S. Multiple sclerosis shrinks intralaminar, and enlarges extralaminar, brain parenchymal veins. *Neurology* **80**, 145–151 (2013).
- Müller, K. *et al.* Detailing intra-lesional venous lumen shrinking in multiple sclerosis investigated by sFLAIR MRI at 7-T. *J. Neurol.* **261**, 2032–2036 (2014).
- Dal-Bianco, A. *et al.* Veins in plaques of multiple sclerosis patients — a longitudinal magnetic resonance imaging study at 7 Tesla. *Eur. Radiol.* **25**, 2913–2920 (2015).
- George, I. C. *et al.* Clinical 3-tesla FLAIR* MRI improves diagnostic accuracy in multiple sclerosis. *Mult. Scler.* <http://dx.doi.org/10.1177/1352458515624975> (2016).
- Lummel, N. *et al.* Presence of a central vein within white matter lesions on susceptibility weighted imaging: a specific finding for multiple sclerosis? *Neuroradiology* **53**, 311–317 (2011).
- Kilsdonk, I. D. *et al.* Improved differentiation between MS and vascular brain lesions using FLAIR* at 7 Tesla. *Eur. Radiol.* **24**, 841–849 (2014).
- Solomon, A. J. *et al.* 'Central vessel sign' on 3T FLAIR* MRI for the differentiation of multiple sclerosis from migraine. *Ann. Clin. Transl. Neurol.* **3**, 82–87 (2015).
- Wuerfel, J. *et al.* Lesion morphology at 7 Tesla MRI differentiates Susac syndrome from multiple sclerosis. *Mult. Scler.* **18**, 1592–1599 (2012).
- Massacesi, L. *et al.* Active and passively induced experimental autoimmune encephalomyelitis in common marmosets: a new model for multiple sclerosis. *Ann. Neurol.* **37**, 519–530 (1995).
- Genain, C. P. & Hauser, S. L. Experimental allergic encephalomyelitis in the New World monkey *Callithrix jacchus*. *Immunol. Rev.* **183**, 159–172 (2001).
- Jagessar, S. A. *et al.* Autoimmunity against myelin oligodendrocyte glycoprotein is dispensable for the initiation although essential for the progression of chronic encephalomyelitis in common marmosets. *J. Neuropathol. Exp. Neurol.* **67**, 326–340 (2008).
- Jagessar, S. A., Dijkman, K., Dunham, J., t Hart, B. A. & Kap, Y. S. Experimental autoimmune encephalomyelitis in marmosets. *Methods Mol. Biol.* **1304**, 171–186 (2016).
- Hart, B. A. *et al.* Histopathological characterization of magnetic resonance imaging-detectable brain white matter lesions in a primate model of multiple sclerosis: a correlative study in the experimental autoimmune encephalomyelitis model in common marmosets (*Callithrix jacchus*). *Am. J. Pathol.* **153**, 649–663 (1998).
- t Hart, B. A. *et al.* Modelling of multiple sclerosis: lessons learned in a non-human primate. *Lancet Neurol.* **3**, 588–597 (2004).
- Blezer, E. L., Bauer, J., Brok, H. P., Nicolay, K. & t Hart, B. A. Quantitative MRI-pathology correlations of brain white matter lesions developing in a non-human primate model of multiple sclerosis. *NMR Biomed.* **20**, 90–103 (2007).
- Jordan, E. K. *et al.* Serial MR imaging of experimental autoimmune encephalomyelitis induced by human white matter or by chimeric myelin-basic and proteolipid protein in the common marmoset. *AJNR Am. J. Neuroradiol.* **20**, 965–976 (1999).
- Diem, R. *et al.* Autoimmune optic neuritis in the common marmoset monkey: comparison of visual evoked potentials with MRI and histopathology. *Invest. Ophthalmol. Vis. Sci.* **49**, 3707–3714 (2008).
- Boretius, S. *et al.* Monitoring of EAE onset and progression in the common marmoset monkey by sequential high-resolution 3D MRI. *NMR Biomed.* **19**, 41–49 (2006).
- Gaitan, M. I. *et al.* Perivascular brain lesions in a primate multiple sclerosis model at 7-tesla magnetic resonance imaging. *Mult. Scler.* **20**, 64–71 (2014).
- Henderson, A. P., Barnett, M. H., Parratt, J. D. & Prineas, J. W. Multiple sclerosis: distribution of inflammatory cells in newly forming lesions. *Ann. Neurol.* **66**, 739–753 (2009).
- Prineas, J. W. & Parratt, J. D. Oligodendrocytes and the early multiple sclerosis lesion. *Ann. Neurol.* **72**, 18–31 (2012).
- Lucchinetti, C. *et al.* Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann. Neurol.* **47**, 707–717 (2000).
- Singh, S. *et al.* Microglial nodules in early multiple sclerosis white matter are associated with degenerating axons. *Acta Neuropathol.* **125**, 595–608 (2013).
- Peferoen, L. A. *et al.* Activation status of human microglia is dependent on lesion formation stage and remyelination in multiple sclerosis. *J. Neuropathol. Exp. Neurol.* **74**, 48–63 (2015).
- Nikic, I. *et al.* A reversible form of axon damage in experimental autoimmune encephalomyelitis and multiple sclerosis. *Nat. Med.* **17**, 495–499 (2011).

44. Alvarez, J. I. *et al.* Focal disturbances in the blood–brain barrier are associated with formation of neuroinflammatory lesions. *Neurobiol. Dis.* **74**, 14–24 (2014).
45. Filippi, M., Rocca, M. A., Martino, G., Horsfield, M. A. & Comi, G. Magnetization transfer changes in the normal appearing white matter precede the appearance of enhancing lesions in patients with multiple sclerosis. *Ann. Neurol.* **43**, 809–814 (1998).
46. Goodkin, D. E. *et al.* A serial study of new MS lesions and the white matter from which they arise. *Neurology* **51**, 1689–1697 (1998).
47. Werring, D. J. *et al.* The pathogenesis of lesions and normal-appearing white matter changes in multiple sclerosis: a serial diffusion MRI study. *Brain* **123**, 1667–1676 (2000).
48. Eisele, P. *et al.* Reduced diffusion in a subset of acute MS lesions: a serial multiparametric MRI study. *AJNR Am. J. Neuroradiol.* **33**, 1369–1373 (2012).
49. Tartaglia, M. C. *et al.* Choline is increased in pre-lesional normal appearing white matter in multiple sclerosis. *J. Neurol.* **249**, 1382–1390 (2002).
50. Wuerfel, J. *et al.* Changes in cerebral perfusion precede plaque formation in multiple sclerosis: a longitudinal perfusion MRI study. *Brain* **127**, 111–119 (2004).
51. Wiggermann, V. *et al.* Magnetic resonance frequency shifts during acute MS lesion formation. *Neurology* **81**, 211–218 (2013).
52. Guttman, C. R. *et al.* Multiple sclerosis lesion formation and early evolution revisited: a weekly high-resolution magnetic resonance imaging study. *Mult. Scler.* **22**, 761–769 (2016).
53. Obermeier, B., Daneman, R. & Ransohoff, R. M. Development, maintenance and disruption of the blood–brain barrier. *Nat. Med.* **19**, 1584–1596 (2013).
54. Vos, C. M. *et al.* Blood–brain barrier alterations in both focal and diffuse abnormalities on postmortem MRI in multiple sclerosis. *Neurobiol. Dis.* **20**, 953–960 (2005).
55. Gay, D. & Esiri, M. Blood–brain barrier damage in acute multiple sclerosis plaques: an immunocytological study. *Brain* **114**, 557–572 (1991).
56. Agrawal, S. M. *et al.* Extracellular matrix metalloproteinase inducer shows active perivascular cuffs in multiple sclerosis. *Brain* **136**, 1760–1777 (2013).
57. Davalos, D. *et al.* Fibrinogen-induced perivascular microglial clustering is required for the development of axonal damage in neuroinflammation. *Nat. Commun.* **3**, 1227 (2012).
58. Ryu, J. K. *et al.* Blood coagulation protein fibrinogen promotes autoimmunity and demyelination via chemokine release and antigen presentation. *Nat. Commun.* **6**, 8164 (2015).
59. McFarland, H. F. *et al.* Using gadolinium-enhanced magnetic resonance imaging lesions to monitor disease activity in multiple sclerosis. *Ann. Neurol.* **32**, 758–766 (1992).
60. Frank, J. A. *et al.* Serial contrast-enhanced magnetic resonance imaging in patients with early relapsing–remitting multiple sclerosis: implications for treatment trials. *Ann. Neurol.* **36**, S86–S90 (1994).
61. Gaitan, M. I. *et al.* Evolution of the blood–brain barrier in newly forming multiple sclerosis lesions. *Ann. Neurol.* **70**, 22–29 (2011).
62. Gaitan, M. I., Sati, P., Inati, S. J. & Reich, D. S. Initial investigation of the blood–brain barrier in MS lesions at 7 tesla. *Mult. Scler.* **19**, 1068–1073 (2013).
63. Lassmann, H., Raine, C. S., Antel, J. & Prineas, J. W. Immunopathology of multiple sclerosis: report on an international meeting held at the Institute of Neurology of the University of Vienna. *J. Neuroimmunol.* **86**, 213–217 (1998).
64. van der Valk, P. & De Groot, C. J. Staging of multiple sclerosis (MS) lesions: pathology of the time frame of MS. *Neuropathol. Appl. Neurobiol.* **26**, 2–10 (2000).
65. Yamasaki, R. *et al.* Differential roles of microglia and monocytes in the inflamed central nervous system. *J. Exp. Med.* **211**, 1533–1549 (2014).
66. Butovsky, O. *et al.* Identification of a unique TGF- β -dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17**, 131–143 (2014).
67. Miron, V. E. *et al.* M2 microglia and macrophages drive oligodendrocyte differentiation during CNS remyelination. *Nat. Neurosci.* **16**, 1211–1218 (2013).
68. Lassmann, H., van Horssen, J. & Mahad, D. Progressive multiple sclerosis: pathology and pathogenesis. *Nat. Rev. Neurol.* **8**, 647–656 (2012).
69. Hametner, S. *et al.* Iron and neurodegeneration in the multiple sclerosis brain. *Ann. Neurol.* **74**, 848–861 (2013).
70. Stephenson, E., Nathoo, N., Mahjoub, Y., Dunn, J. F. & Yong, V. W. Iron in multiple sclerosis: roles in neurodegeneration and repair. *Nat. Rev. Neurol.* **10**, 459–468 (2014).
71. Trapp, B. D. *et al.* Axonal transection in the lesions of multiple sclerosis. *N. Engl. J. Med.* **338**, 278–285 (1998).
72. Dutta, R. *et al.* Mitochondrial dysfunction as a cause of axonal degeneration in multiple sclerosis patients. *Ann. Neurol.* **59**, 478–489 (2006).
73. Laule, C. *et al.* Myelin water imaging in multiple sclerosis: quantitative correlations with histopathology. *Mult. Scler.* **12**, 747–753 (2006).
74. Levesque, I. R. *et al.* Reproducibility of quantitative magnetization-transfer imaging parameters from repeated measurements. *Magn. Reson. Med.* **64**, 391–400 (2010).
75. Sati, P. *et al.* Micro-compartment specific T₂* relaxation in the brain. *NeuroImage* **77**, 268–278 (2013).
76. Alonso-Ortiz, E., Levesque, I. R. & Pike, G. B. MRI-based myelin water imaging: a technical review. *Magn. Reson. Med.* **73**, 70–81 (2015).
77. Yablonsky, D. A., Luo, J., Sukstanskii, A. L., Iyer, A. & Cross, A. H. Biophysical mechanisms of MRI signal frequency contrast in multiple sclerosis. *Proc. Natl Acad. Sci. USA* **109**, 14212–14217 (2012).
78. Mehta, V. *et al.* Iron is a sensitive biomarker for inflammation in multiple sclerosis lesions. *PLoS ONE* **8**, e57573 (2013).
79. Chen, W. *et al.* Quantitative susceptibility mapping of multiple sclerosis lesions at various ages. *Radiology* **271**, 183–192 (2014).
80. Schechter, R. & Schwartz, M. CNS sterile injury: just another wound healing? *Trends Mol. Med.* **19**, 135–143 (2013).
81. Prineas, J. W., Barnard, R. O., Kwon, E. E., Sharer, L. R. & Cho, E. S. Multiple sclerosis: remyelination of nascent lesions. *Ann. Neurol.* **33**, 137–151 (1993).
82. Raine, C. S. & Wu, E. Multiple sclerosis: remyelination in acute lesions. *J. Neuropathol. Exp. Neurol.* **52**, 199–204 (1993).
83. Patrikios, P. *et al.* Remyelination is extensive in a subset of multiple sclerosis patients. *Brain* **129**, 3165–3172 (2006).
84. Albert, M., Antel, J., Bruck, W. & Stadelmann, C. Extensive cortical remyelination in patients with chronic multiple sclerosis. *Brain Pathol.* **17**, 129–138 (2007).
85. Bramow, S. *et al.* Demyelination versus remyelination in progressive multiple sclerosis. *Brain* **133**, 2983–2998 (2010).
86. Chang, A., Tourtellotte, W. W., Rudick, R. & Trapp, B. D. Remyelinating oligodendrocytes in chronic lesions of multiple sclerosis. *N. Engl. J. Med.* **346**, 165–173 (2002).
87. Kuhlmann, T. *et al.* Differentiation block of oligodendroglial progenitor cells as a cause for remyelination failure in chronic multiple sclerosis. *Brain* **131**, 1749–1758 (2008).
88. Goldschmidt, T., Antel, J., König, F. B., Bruck, W. & Kuhlmann, T. Remyelination capacity of the MS brain decreases with disease chronicity. *Neurology* **72**, 1914–1921 (2009).
89. Franklin, R. J. & Gallo, V. The translational biology of remyelination: past, present, and future. *Glia* **62**, 1905–1915 (2014).
90. van Walderveen, M. A. *et al.* Histopathologic correlate of hypointense lesions on T1-weighted spin-echo MRI in multiple sclerosis. *Neurology* **50**, 1282–1288 (1998).
91. van Waesberghe, J. H. *et al.* Axonal loss in multiple sclerosis lesions: magnetic resonance imaging insights into substrates of disability. *Ann. Neurol.* **46**, 747–754 (1999).
92. Chen, J. T. *et al.* Magnetization transfer ratio evolution with demyelination and remyelination in multiple sclerosis lesions. *Ann. Neurol.* **63**, 254–262 (2008).
93. Reich, D. S. *et al.* Sample-size calculations for short-term proof-of-concept studies of tissue protection and repair in multiple sclerosis lesions via conventional clinical imaging. *Mult. Scler.* **21**, 1693–1704 (2015).
94. Lassmann, H. Review: the architecture of inflammatory demyelinating lesions: implications for studies on pathogenesis. *Neuropathol. Appl. Neurobiol.* **37**, 698–710 (2011).
95. Frischer, J. M. *et al.* Clinical and pathological insights into the dynamic nature of the white matter multiple sclerosis plaque. *Ann. Neurol.* **78**, 710–721 (2015).
96. Pitt, D. *et al.* Imaging cortical lesions in multiple sclerosis with ultra-high-field magnetic resonance imaging. *Arch. Neurol.* **67**, 812–818 (2010).
97. Bagnato, F. *et al.* Tracking iron in multiple sclerosis: a combined imaging and histopathological study at 7 Tesla. *Brain* **134**, 3602–3615 (2011).
98. Bian, W. *et al.* A serial in vivo 7T magnetic resonance phase imaging study of white matter lesions in multiple sclerosis. *Mult. Scler.* **19**, 69–75 (2012).
99. Yao, B. *et al.* Chronic multiple sclerosis lesions: characterization with high-field-strength MR imaging. *Radiology* **262**, 206–215 (2012).
100. Hagemeyer, J. *et al.* Iron deposition in multiple sclerosis lesions measured by susceptibility-weighted imaging filtered phase: a case control study. *J. Magn. Reson. Imaging* **36**, 73–83 (2012).
101. Absinta, M. *et al.* Poor outcome in MS lesions with persistent 7-tesla phase rim. *J. Clin. Invest.* <http://dx.doi.org/10.1172/JCI86198> (2016).
102. Vellinga, M. M. *et al.* Pluriformity of inflammation in multiple sclerosis shown by ultra-small iron oxide particle enhancement. *Brain* **131**, 800–807 (2008).
103. Tourdias, T. *et al.* Assessment of disease activity in multiple sclerosis phenotypes with combined gadolinium- and superparamagnetic iron oxide-enhanced MR imaging. *Radiology* **264**, 225–233 (2012).
104. Maarouf, A. *et al.* Ultra-small superparamagnetic iron oxide enhancement is associated with higher loss of brain tissue structure in clinically isolated syndrome. *Mult. Scler.* <http://dx.doi.org/10.1177/1352458515607649> (2015).
105. Doussot, V. *et al.* Comparison of ultrasmall particles of iron oxide (USPIO)-enhanced T2-weighted, conventional T2-weighted, and gadolinium-enhanced T1-weighted MR images in rats with experimental autoimmune encephalomyelitis. *AJNR Am. J. Neuroradiol.* **20**, 223–227 (1999).
106. Berger, C., Hiestand, P., Kandler-Baumann, D., Rudin, M. & Rausch, M. Analysis of lesion development during acute inflammation and remission in a rat model of experimental autoimmune encephalomyelitis by visualization of macrophage infiltration, demyelination and blood–brain barrier damage. *NMR Biomed.* **19**, 101–107 (2006).
107. Baeten, K. *et al.* Visualisation of the kinetics of macrophage infiltration during experimental autoimmune encephalomyelitis by magnetic resonance imaging. *J. Neuroimmunol.* **195**, 1–6 (2008).
108. Lucchinetti, C. F., Bruck, W., Rodriguez, M. & Lassmann, H. Distinct patterns of multiple sclerosis pathology indicates heterogeneity on pathogenesis. *Brain Pathol.* **6**, 259–274 (1996).
109. Metz, I. *et al.* Pathologic heterogeneity persists in early active multiple sclerosis lesions. *Ann. Neurol.* **75**, 728–738 (2014).
110. Popescu, B. F., Bunyan, R. F., Parisi, J. E., Ransohoff, R. M. & Lucchinetti, C. F. A case of multiple sclerosis presenting with inflammatory cortical demyelination. *Neurology* **76**, 1705–1710 (2011).
111. Kutzelnigg, A. *et al.* Cortical demyelination and diffuse white matter injury in multiple sclerosis. *Brain* **128**, 2705–2712 (2005).
112. Peterson, J. W., Bo, L., Mork, S., Chang, A. & Trapp, B. D. Transected neurites, apoptotic neurons, and reduced inflammation in cortical multiple sclerosis lesions. *Ann. Neurol.* **50**, 389–400 (2001).
113. Bo, L., Vedeler, C. A., Nyland, H., Trapp, B. D. & Mork, S. J. Intracortical multiple sclerosis lesions are not associated with increased lymphocyte infiltration. *Mult. Scler.* **9**, 323–331 (2003).
114. Kooi, E. J., Strijbis, E. M., van der Valk, P. & Geurts, J. J. Heterogeneity of cortical lesions in multiple sclerosis: clinical and pathological implications. *Neurology* **79**, 1369–1376 (2012).
115. Chang, A. *et al.* Cortical remyelination: a new target for repair therapies in multiple sclerosis. *Ann. Neurol.* **72**, 918–926 (2012).
116. Kidd, D. *et al.* Cortical lesions in multiple sclerosis. *Brain* **122**, 17–26 (1999).
117. Fischer, M. T. *et al.* Disease-specific molecular events in cortical multiple sclerosis lesions. *Brain* **136**, 1799–1815 (2013).
118. Magliozzi, R. *et al.* Meningeal B-cell follicles in secondary progressive multiple sclerosis associate with early onset of disease and severe cortical pathology. *Brain* **130**, 1089–1104 (2007).
119. Magliozzi, R. *et al.* A Gradient of neuronal loss and meningeal inflammation in multiple sclerosis. *Ann. Neurol.* **68**, 477–493 (2010).

120. Howell, O. W. *et al.* Meningeal inflammation is widespread and linked to cortical pathology in multiple sclerosis. *Brain* **134**, 2755–2771 (2011).
121. Choi, S. R. *et al.* Meningeal inflammation plays a role in the pathology of primary progressive multiple sclerosis. *Brain* **135**, 2925–2937 (2012).
122. Kuerten, S. *et al.* Tertiary lymphoid organ development coincides with determinant spreading of the myelin-specific T cell response. *Acta Neuropathol.* **124**, 861–873 (2012).
123. Magliozzi, R. *et al.* B-cell enrichment and Epstein–Barr virus infection in inflammatory cortical lesions in secondary progressive multiple sclerosis. *J. Neuropathol. Exp. Neurol.* **72**, 29–41 (2013).
124. Bo, L., Geurts, J. J., van der Valk, P., Polman, C. & Barkhof, F. Lack of correlation between cortical demyelination and white matter pathologic changes in multiple sclerosis. *Arch. Neurol.* **64**, 76–80 (2007).
125. Calabrese, M. *et al.* Exploring the origins of grey matter damage in multiple sclerosis. *Nat. Rev. Neurosci.* **16**, 147–158 (2015).
126. Geurts, J. J. *et al.* Cortical lesions in multiple sclerosis: combined postmortem MR imaging and histopathology. *AJNR Am. J. Neuroradiol.* **26**, 572–577 (2005).
127. Schmierer, K. *et al.* High field (9.4 Tesla) magnetic resonance imaging of cortical grey matter lesions in multiple sclerosis. *Brain* **133**, 858–867 (2010).
128. Seewann, A. *et al.* Postmortem verification of MS cortical lesion detection with 3D DIR. *Neurology* **78**, 302–308 (2012).
129. Sethi, V. *et al.* Improved detection of cortical MS lesions with phase-sensitive inversion recovery MRI. *J. Neurol. Neurosurg. Psychiatry* **83**, 877–882 (2012).
130. Mainiero, C. *et al.* *In vivo* imaging of cortical pathology in multiple sclerosis using ultra-high field MRI. *Neurology* **73**, 941–948 (2009).
131. Nielsen, A. S. *et al.* Contribution of cortical lesion subtypes at 7T MRI to physical and cognitive performance in MS. *Neurology* **81**, 641–649 (2013).
132. Mainiero, C. *et al.* A gradient in cortical pathology in multiple sclerosis by *in vivo* quantitative 7 T imaging. *Brain* **138**, 932–945 (2015).
133. Abdel-Fahim, R. *et al.* Improved detection of focal cortical lesions using 7 T magnetisation transfer imaging in patients with multiple sclerosis. *Mult. Scler. Relat. Disord.* **3**, 258–265 (2014).
134. Jonkman, L. E. *et al.* Ultra-high field MTR and qR2* differentiates subpial cortical lesions from normal-appearing gray matter in multiple sclerosis. *Mult. Scler.* <http://dx.doi.org/10.1177/1352458515620499> (2015).
135. Calabrese, M. *et al.* Cortical lesions in primary progressive multiple sclerosis: a 2-year longitudinal MR study. *Neurology* **72**, 1330–1336 (2009).
136. Calabrese, M. *et al.* Cortical lesions and atrophy associated with cognitive impairment in relapsing–remitting multiple sclerosis. *Arch. Neurol.* **66**, 1144–1150 (2009).
137. Roosendaal, S. D. *et al.* Accumulation of cortical lesions in MS: relation with cognitive impairment. *Mult. Scler.* **15**, 708–714 (2009).
138. Calabrese, M. *et al.* A 3-year magnetic resonance imaging study of cortical lesions in relapse-onset multiple sclerosis. *Ann. Neurol.* **67**, 376–383 (2010).
139. Calabrese, M. *et al.* Cortical lesion load associates with progression of disability in multiple sclerosis. *Brain* **135**, 2952–2961 (2012).
140. Absinta, M. *et al.* Gadolinium-based MRI characterization of leptomeningeal inflammation in multiple sclerosis. *Neurology* **85**, 18–28 (2015).
141. Pomeroy, I. M., Matthews, P. M., Frank, J. A., Jordan, E. K. & Esiri, M. M. Demyelinated neocortical lesions in marmoset autoimmune encephalomyelitis mimic those in multiple sclerosis. *Brain* **128**, 2713–2721 (2005).
142. Merkler, D. *et al.* Differential macrophage/microglia activation in neocortical EAE lesions in the marmoset monkey. *Brain Pathol.* **16**, 117–123 (2006).
143. Pomeroy, I. M., Jordan, E. K., Frank, J. A., Matthews, P. M. & Esiri, M. M. Diffuse cortical atrophy in a marmoset model of multiple sclerosis. *Neurosci. Lett.* **437**, 121–124 (2008).
144. Pomeroy, I. M., Jordan, E. K., Frank, J. A., Matthews, P. M. & Esiri, M. M. Focal and diffuse cortical degenerative changes in a marmoset model of multiple sclerosis. *Mult. Scler.* **16**, 537–548 (2010).
145. Kramann, N. *et al.* Increased meningeal T and plasma cell infiltration is associated with early subpial cortical demyelination in common marmosets with experimental autoimmune encephalomyelitis. *Brain Pathol.* **25**, 276–286 (2015).

Acknowledgements

The authors acknowledge financial support from the Intramural Research Program of the National Institute of Neurological Disorders and Stroke, USA. D.S.R. has received research support from Vertex Pharmaceuticals and the Myelin Repair Foundation. This funding is not a competing interest in the context of the current article.

Author contributions

All authors researched data for the article, made substantial contributions to discussions of the content, wrote the article, and reviewed and/or edited the manuscript before submission.

Competing interests statement

The authors declare no competing interests.

SUPPLEMENTARY INFORMATION

See online article: S1 (video) | S2 (video) | S3 (video)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

From blood–brain barrier to blood–brain interface: new opportunities for CNS drug delivery

William A. Banks

Abstract | One of the biggest challenges in the development of therapeutics for central nervous system (CNS) disorders is achieving sufficient blood–brain barrier (BBB) penetration. Research in the past few decades has revealed that the BBB is not only a substantial barrier for drug delivery to the CNS but also a complex, dynamic interface that adapts to the needs of the CNS, responds to physiological changes, and is affected by and can even promote disease. This complexity confounds simple strategies for drug delivery to the CNS, but provides a wealth of opportunities and approaches for drug development. Here, I review some of the most important areas that have recently redefined the BBB and discuss how they can be applied to the development of CNS therapeutics.

Blood–brain barrier (BBB). The modified capillary bed of the brain; can be conceptualized as those processes that, taken together, control the exchange of substances between the blood and the fluids (cerebrospinal fluid and brain interstitial fluid) of the central nervous system (CNS).

Developing therapeutics for brain diseases is a major challenge, and a particularly formidable aspect of that challenge is the blood–brain barrier (BBB). Named early in the twentieth century for its ability to prevent the uncontrolled leakage of substances from the blood into the brain, the BBB has emerged as a complex, dynamic, adaptable interface that controls the exchange of substances between the central nervous system (CNS) and the blood (BOX 1). The cells that make up the structure of the BBB (FIG. 1) communicate with the other cells of the CNS, adapting their behaviour to serve the needs of the CNS, responding to pathological conditions, and in some cases participating in the onset, maintenance or progression of disease. On the one hand, this complexity in barrier function explains much of the difficulty in developing drugs that can cross the BBB, but on the other hand it offers unique and diverse opportunities for drug development.

Here, I review the work in five areas from the past 30 years that has the potential to transform approaches to CNS drug delivery, first considering new work and some of the novel approaches it presents, and then considering how this work modifies classical approaches. Although there have been advances in BBB research in other areas, those selected for discussion in this Review are considered to have most relevance to drug development. I then consider specific diseases in which the BBB itself is an active player and how the BBB can be targeted therapeutically. Finally, I discuss future directions for CNS drug delivery that combine new and classical approaches.

From ‘barrier’ to ‘interface’

The past 30 years have seen a great deal of research on the BBB and the fields related to it. This work has expanded our understanding of how the BBB functions, how it interacts with its environment, how it can be affected by disease, and how it can predispose to or even cause disease. Some of the work has come from an expansion of traditional areas of BBB research, but much more comes from integration with fields that, 30 years ago, would have been thought to have little or no connection with the BBB. This work can be categorized into the following five areas: the neurovascular unit (NVU), transporters, barrier cell secretions, polarization, and adaptations and modifications of BBB functions. As discussed below, these advances suggest new strategies for CNS drug delivery and ways to improve on traditional CNS drug delivery approaches, and also demonstrate several diseases in which the BBB itself should be the therapeutic target.

The neurovascular unit. The anatomical locations of barrier functions in the CNS are complex. The vascular BBB, the spinal cord barrier and the barriers of the cranial nerves are formed by endothelial cells, and the blood–cerebrospinal fluid (CSF) barrier at the choroid plexus by modified epithelial (ependymal) cells¹. Circumventricular organs — small regions of the brain with vasculature that does not form a tight BBB — are separated from the rest of the brain by a barrier formed by tanycytes and from the adjacent CSF by ependymal cells. The blood–retinal barrier consists of an inner

Veterans Affairs Puget Sound Health Care System, Geriatrics Research Education and Clinical Center and Department of Medicine, University of Washington School of Medicine, Division of Gerontology and Geriatric Medicine, 1660 South Columbian Way, Seattle, Washington 98108, USA. wabanks1@uw.edu

doi:10.1038/nrd.2015.21
Published online 22 Jan 2016

Box 1 | The first 100 years of BBB research

Experimental evidence for a barrier to the movement of solutes from blood into the tissue bed of the central nervous system (CNS) dates back to the 1880s. The most widely known experiments are those of Paul Ehrlich, who demonstrated that many dyes did not stain the CNS after their peripheral injection. The dyes bound tightly to albumin, leaving almost no unbound dye available for passage into the CNS. These studies, therefore, demonstrated the impermeability of the blood–brain barrier (BBB) to large circulating proteins. By contrast, studies from the 1920s to 1940s showed that cerebrospinal fluid (CSF) levels of glucose and inorganic compounds, including electrolytes, mirrored the levels found in blood^{169,170}. These CSF/serum ratios differed among substances and were altered in disease. Thus, the barrier phenomenon was not absolute, but somehow differentiated among substances and responded to pathophysiological conditions.

By the late 1950s, one basis of differentiation among substances was well established: the BBB often behaved like a lipid membrane¹⁷¹. Small, lipid-soluble substances could readily enter the CNS, whereas more water-soluble drugs, those that were highly charged, or those that bound tightly to serum proteins could not. However, many researchers doubted whether a physical barrier existed or could otherwise explain the many complex phenomena attributed to the barrier¹⁷², until the late 1960s. Then, ultrastructural studies showed that the capillary bed of the brain possessed tight junctions between brain endothelial cells (BECs) as well as a virtual absence of macropinocytotic vesicles and intracellular fenestrae¹⁷³. Tight junctions were also found at the choroid plexus, thus explaining the presence of the blood–CSF barrier. These studies showed that at least two parallel barriers existed: a blood–CSF barrier located at the choroid plexus, and a vascular BBB located at the arteriole–capillary–venule level.

Transporters provided another mechanism by which the barrier could differentiate among substances, explaining how glucose, electrolytes, vitamins and some peptides and regulatory proteins can cross the barrier at rates that are 10–100 times faster than those predicted by the lipid membrane model.

In summary, the first 100 years of study of the BBB (1880s–1980s), established some central, somewhat paradoxical features of the BBB. The next 30 years would see barrier research explode, adding many new aspects to the understanding of the barrier phenomenon.

Neurovascular unit

(NVU). For the purposes of this Review, this refers to the concept that the cells forming the BBB are in communication with other cells of the central nervous system (CNS) and, by extension, with the circulating immune cells, and with the peripheral tissues via blood-borne secretions.

Transporters

Proteins that provide a mechanism by which substances can be carried from one side of the blood–brain barrier (BBB) to the other, thus greatly increasing (for blood-to-brain transporters) or greatly decreasing (for brain-to-blood transporters) the central nervous system (CNS) uptake of a substance in comparison with that predicted based on its physicochemical characteristics.

barrier of endothelial cells and an outer barrier involving the retinal pigment epithelium. Many of these barriers have been poorly studied, but it is clear that each of the barriers is adapted to the special needs of the tissues it serves, responds to disease states, has intercellular tight junctions as a central feature to its barrier function, and directly interacts with the cells in its microenvironment.

The vascular BBB is the best studied regarding this last phenomenon of intercellular interactions: the brain endothelial cells (BECs) and the cells with which they interact are collectively referred to as the NVU (FIG. 2). The concept of the NVU emphasizes that BECs are in intimate and constant crosstalk with astrocytes, microglia, neurons, mast cells and pericytes, as well as circulating immune cells². This communication helps refine the functions of the BBB to serve the needs of the brain and facilitates brain–body communication. The astrocytes and pericytes in particular are responsible for enabling the capillary bed to adopt and maintain barrier characteristics^{3,4}.

Neuroimmune modulators are clearly involved in NVU–BBB crosstalk⁵. Release of cytokines and other immunoactive substances from astrocytes, microglia, leukocytes and even the BECs themselves modify a host of BBB functions, altering BBB integrity, BBB transporters, and the permeability of the BBB to pathogens and circulating immune cells. Increasingly elaborate examples of NVU communications are being elucidated.

For example, in mouse models of Alzheimer disease, astrocyte-derived apolipoprotein E (APOE) modulates a peptidyl-prolyl *cis-trans* isomerase A (PPIase A; also known as cyclophilin A)–matrix metalloproteinase (MMP) pathway in pericytes, activation of which results in BBB disruption⁶. As discussed below, this and other NVU interactions offer unique therapeutic targets in Alzheimer disease, diabetes and other conditions.

Transporters. Although transporters are a classical characteristic of the BBB, so much has been learned about them during the past 30 years as to redefine them for the purposes of CNS drug delivery. The presence of BBB transporters explains how water-soluble molecules such as glucose and amino acids that are needed in abundance by the CNS can cross the BBB at rates that are 10–100 times faster than would be predicted from their physicochemical characteristics⁷. Transporters also mediate the CNS uptake of large molecules such as regulatory proteins, which cannot cross the BBB by transendothelial diffusion⁸. The presence of CNS-to-blood transport (efflux) systems explains why many substances enter or accumulate in the CNS at rates lower than would be predicted based on their physicochemical characteristics. Active transport refers to an energy-requiring system and can transport a substance against its concentration gradient, whereas facilitated diffusion does not require energy and transports its ligands down a concentration gradient, from higher to lower concentrations.

Blood-to-brain transporters. Transporters exist for a surprising array of molecules, including cytokines⁹ and oligonucleotide analogues¹⁰. Phosphorothioate oligonucleotides are rapidly sequestered by and transported across the BBB. The transporter seems promiscuous towards phosphorothioate oligonucleotides, as it has been shown to transport into the brain every phosphorothioate oligonucleotide tested so far. The transporter, when combined with the highly favourable pharmacokinetics of phosphorothioate oligonucleotides, enables the delivery of these molecules to the brain in therapeutically effective amounts. These molecules have shown therapeutic benefits in animal models of multiple sclerosis¹¹, Alzheimer disease^{10,12,13} and stroke¹⁴. For example, peripheral administration of an antisense oligonucleotide directed against amyloid precursor protein decreased levels of amyloid- β_{40} ($A\beta_{40}$), $A\beta_{42}$ and amyloid precursor protein in the brain, improved cognitive deficits, decreased oxidative stress, restored deficits in the brain-to-blood efflux of $A\beta$ and in bulk flow, normalized brain levels of low-density lipoprotein receptor-related protein 1 (LRP1), and reduced neuroinflammatory markers in animal models of Alzheimer disease^{12,15,16}.

The existence of saturable transporters and their unique characteristics offer substantial untapped potential for drug development. The endogenous ligands of transporters are seldom useful as therapeutics because of their unfavourable pharmacokinetic properties. Brain-derived neurotrophic factor, for example, crosses the BBB rapidly^{17,18}, but its short half-life in blood limits its CNS effects. One strategy is to develop drug analogues of

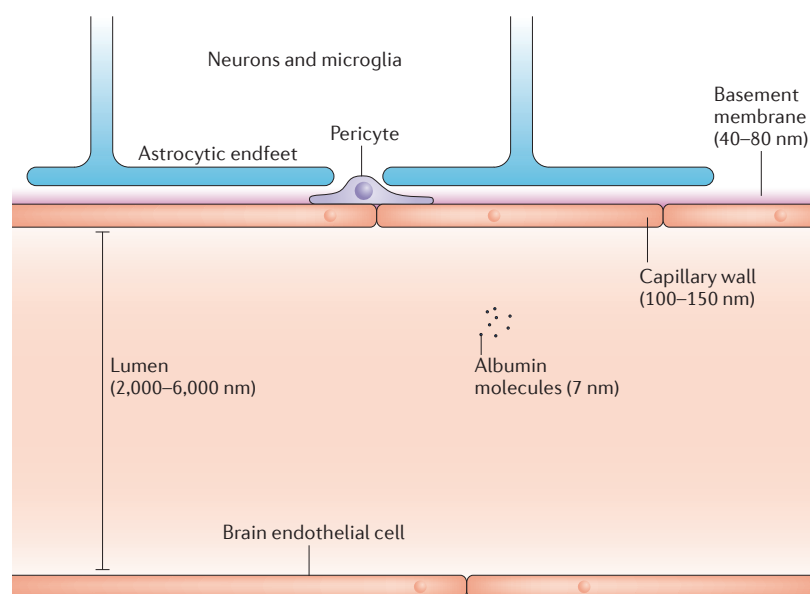


Figure 1 | Proportions of the blood–brain barrier. Schematic with dimensions of the vascular blood–brain barrier (BBB), including the capillary lumen and walls, and the basement membrane, relatively scaled. Note the albumin molecules with a diameter of 7 nm. Other components of the neurovascular unit are indicated, but not scaled.

endogenous ligands that can use the BBB transporters to enter the CNS. However, this does not seem to have been a widely exploited strategy, and few CNS drugs are known to use transporters to cross the BBB. Examples include: L-dopa, melphalan, baclofen, thyroid hormones and gabapentin, which use transporters for neutral amino acids; valproic acid, which probably uses a monocarboxylate transporter; and acetylcholinesterase inhibitors, verapamil, levofloxacin and cephaloridine, which use an organic cation–carnitine transporter. A major challenge is to develop analogues that are resistant to the enzymatic and clearance mechanisms of the ligand but are still able to bind with high affinity to the ligand's BBB transporter and CNS receptors.

A therapeutic antagonist effect can be induced by blocking a substance from accessing its BBB transporter, thus preventing it from reaching the CNS. Classic examples are the manipulation of amino acid uptake into the brain by the large neutral amino acid (LNAA) transporter. LNAAs have many roles in the brain, including acting as precursors for monoamines and neurotoxic substances. By inhibiting the uptake of one LNAA with another LNAA or a subclass of LNAAs, one can affect brain function and disease responses¹⁹. An example is maple syrup urine disease, a condition in which an excess of branched chain amino acids (BCAAs) are taken up into the brain by the LNAA transporter, resulting in neurotoxic effects. Raising blood levels of other LNAAs competitively inhibits the transport of BCAAs into the brain by the LNAA transporter, thus lowering brain levels of BCAAs and reducing neurotoxicity. A similar strategy has been suggested for phenylketonuria, a condition in which an excess of the LNAA phenylalanine produces neurotoxicity. Blockade of the LNAA transporter with BCAAs has been attempted

to reduce brain levels of tyrosine in individuals with bipolar disorder, to reduce brain phenylalanine levels in tardive dyskinesia, and to lower brain levels of glutamate in amyotrophic lateral sclerosis¹⁹. Another example is the use of a pegylated leptin analogue that does not cross the BBB but binds to the BBB leptin transporter, thus blocking the blood-to-brain transport of endogenous leptin²⁰. This results in increased feeding, which suggests that this analogue could be useful in the treatment of anorexia.

Approaches using the ligands of transporters to deliver therapeutic cargoes have been under development for 30 years, and have been previously reviewed²¹. In this case, the ligand is not the therapeutic, but is used as a vehicle to deliver an attached therapeutic. Although increased BBB permeation is the main objective of this approach, it is likely that the pharmacokinetics of the drug, especially its half-life, are also improved²². One version of this approach that has gained much interest is to develop an antibody targeted to a region of the transporter protein that is not directly involved in binding to the endogenous ligand. In this way, it is hoped that the delivery system will not interfere with the transport of the endogenous ligand into the CNS.

Some pathogens gain access to the CNS using a similar strategy of binding to a transporter or receptor at a non-ligand-binding site. For example, the West Nile virus targets Toll-like receptor 3 (TLR3)²³, whereas herpes varicella zoster virus and cell-free HIV-1 target the mannose-6-phosphate receptor (M6PR)^{24,25}. This observation has prompted the idea that the strategies used by pathogens to cross the BBB might be adaptable to drug delivery²⁶. The cellular machinery involved in the internalization of antibodies and pathogens that have bound to a site not used by the endogenous ligand is likely to be related to mechanisms underlying adsorptive endocytosis and adsorptive transcytosis (AE/AT) rather than to the transporter. A low-affinity binding antibody specific for the BEC transferrin receptor has recently been used to target A β production in an animal model of Alzheimer disease^{27,28}. Recent evidence suggests that immunoglobulin G (IgG) antibodies that have only one antigen-binding arm rather than the typical two may be internalized using a transporter rather than via adsorptive endocytic machinery²⁹.

A drug crossing the BBB by transendothelial diffusion will enter all regions of the brain to a similar degree, thus producing off-target effects in the case of pluripotent drugs. The heterogeneous distribution of many of the transporters for biologics offers a strategy to selectively deliver a drug to a targeted region of the brain. Tumour necrosis factor (TNF), for example, is primarily transported into the hypothalamus and occipital cortex³⁰, whereas the interleukin-1 (IL-1) transporter is highly concentrated at the posterior division of the septum³¹. By using a transporter that targets a specific brain region and thus the function of this region, such as a transporter with enriched expression in the hypothalamus for a feeding effect or a hippocampal transporter for a memory effect, the specificity of a drug can be increased and its side effects decreased.

Polarization

A characteristic of the blood–brain barrier (BBB), arising from different characteristics of its abluminal and luminal surfaces, including differing levels of enzymes, glycoproteins, lipid composition and transporters.

Active transport

Transport by an energy-requiring transporter that can move its ligand against a concentration gradient.

Facilitated diffusion

Transport by a non-energy-requiring transporter that moves its ligand down a concentration gradient.

Adsorptive transcytosis

A mechanism by which glycoproteins or highly charged molecules bind to brain endothelial cell (BEC) glycoproteins, inducing vesicles that are routed to the opposite membrane.

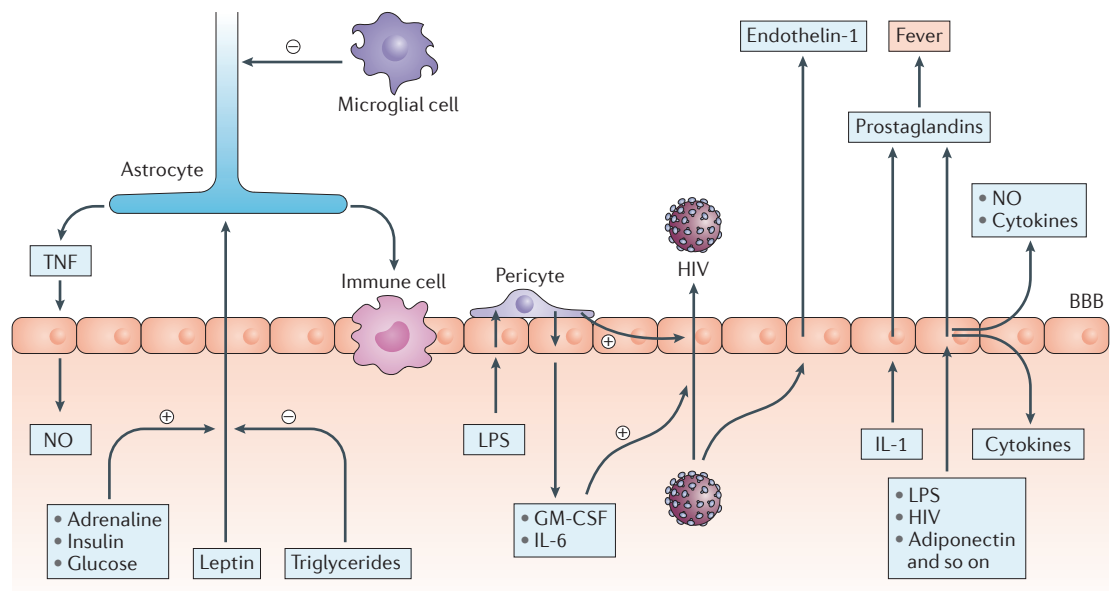


Figure 2 | The neurovascular unit and brain endothelial cell secretions. The ability of brain endothelial cells (BECs) to receive, transfer and transmit signals makes the blood–brain barrier (BBB) an information conduit between the central nervous system (CNS) and periphery. As shown on the left, leptin transport across the BBB is modulated by blood levels of adrenaline, insulin, glucose and triglycerides. Astrocyte control of immune cell trafficking is influenced by leptin and microglia, and astrocytes affect nitric oxide (NO) release from BECs by secreting tumour necrosis factor (TNF). The middle region of the figure shows that lipopolysaccharide (LPS) enhances HIV transport across the BBB — a process that is intensified by pericytes — by inducing the release of granulocyte–macrophage colony-stimulating factor (GM-CSF) and interleukin-6 (IL-6) from BECs; GM-CSF and IL-6 act at the luminal surface of the BEC in an autocrine fashion. HIV induces the release of endothelin-1 from BECs; cerebrospinal fluid (CSF) levels of endothelin-1 correlate with AIDS-related cognitive impairment. The region on the right shows that cytokines, gastrointestinal hormones, viruses and LPS induce the release of substances from BECs that can enter the brain or blood; for example, IL-1 induces the release of prostaglandins from BECs, thus inducing fever.

Brain-to-blood transporters. Efflux transporters can dramatically decrease brain uptake or retention of their substrates just as influx transporters can dramatically increase uptake and retention. Efflux transporters have been discovered for nearly every substance class, including ions, amino acids, peptides and cytokines. Many BBB transporters, including the glucose transporter GLUT1, are facilitated diffusion systems and therefore bidirectional by nature, moving substances down a concentration gradient from a region of higher to lower concentration. Defective efflux transporters can contribute to disease, as illustrated by the decreased brain-to-blood transport of A β peptide. Such efflux ordinarily helps to control brain levels of A β , and loss of efflux contributes to an increasing amyloid burden in the brain. Efflux transporters can also prevent exogenous substances such as plant alkaloids from entering the CNS. Morphine, for example, is a weak substrate for P-glycoprotein (P-gp). P-gp has a broad range of ligands and is a major cause of the inability of many potential therapeutics to enter the brain³².

Inhibition of efflux transporters is exemplified by the inhibition of P-gp in the treatment of seizures with verapamil³³. However, given that so many endogenous and exogenous compounds are ligands for this transporter, long-term inhibition of P-gp would probably result in neurotoxicity and may not, therefore, be feasible.

Inhibition of more targeted efflux systems, however, can improve therapeutic effects. For example, following inhibition of the efflux transporter for pituitary adenylate cyclase-activating polypeptide 27 (PACAP27), a neuroprotective peptide, using an oligophosphorothioate capable of being taken up by the BBB, brain retention of peripherally administered PACAP27 increased by about fourfold and therapeutic outcomes improved in models of stroke and Alzheimer disease¹⁴.

Barrier cell secretions. The discussion above on the NVU emphasizes that secretions from astrocytes, microglia, neurons, pericytes and circulating immune cells have the ability to influence the behaviour of the BBB. However, the BECs and ependymal cells that form the BBB also secrete substances, making communication between the BBB and the other cells of the NVU bidirectional. The characteristics of secretion by barrier cells are similar to those of many other cells in that some secretions are constitutive and others are induced. Many of these secreted substances relate to inflammation, such as prostaglandins, nitric oxide and cytokines^{34–36}.

Some effects of these secretions are mediated in an autocrine fashion; that is, the BEC modifies its own behaviour through its secretions³⁷. For example, activation of the innate immune system induces the release of granulocyte–macrophage colony-stimulating factor

(GM-CSF) and IL-6 from BECs, which then act on the luminal surface of the BECs to enhance the transport of cell-free HIV-1 across the BBB³⁷. This suggests that blood-borne antagonists of GM-CSF and IL-6 could retard HIV-1 entry into the brain and thereby help to reduce the neurological complications associated with HIV/AIDS. Other effects of barrier secretions are mediated indirectly, and are dependent on crosstalk with the other members of the NVU^{38,39}.

Polarization. The cells that form the BBB are polarized in that they have an abluminal surface that faces towards the CNS and a luminal surface that faces towards the blood. Like all cell membranes, those of the barrier cells are lipid bilayers with an inner leaflet in contact with the cytoplasm and an outer leaflet in contact with the interstitial fluid. The ubiquitous tight junctions between cells act like fences, impeding the exchange of lipids and proteins between the luminal and the abluminal outer leaflets^{40,41}. Thus, the lipids and proteins that constitute the leaflets and, by extension, the functions of those leaflets differ between the luminal side and the abluminal side of the BBB.

This polarization increases the complexity of the BBB and its utility to the CNS. The most obvious example is that of transporters: with an ability to restrict transporters to the luminal or abluminal side, unidirectional blood-to-brain or brain-to-blood transport is possible.

Secretions can also be polarized. For example, one *in vitro* study found that BECs secreted IL-6, GM-CSF and TNF from both the luminal and abluminal membranes, but secretions from the luminal membrane were much greater than those from the abluminal membrane⁴². Polarized secretion can also support a form of communication across the BBB, linking the CNS and peripheral tissues via the BBB⁴². Thus, an immune activator on one side of the BBB can induce the release of an immune substance from the other side, thereby forming a unique axis of communication between the peripheral and CNS immune systems. This provides a theoretical therapeutic approach in that a stimulus acting at the luminal side of the BBB could induce it to secrete a desired substance from its abluminal side or, by contrast, a luminal substance could act to block a constitutive abluminal secretion. As an example, peripheral administration of lipopolysaccharide (LPS) induces a biphasic fever by inducing BECs to release prostaglandin E2 (REF. 43). As LPS does not cross the BBB⁴⁴, this means that it is acting on the luminal surface of the BBB to induce the release of prostaglandin E2 from the abluminal surface. As another example, indomethacin, which poorly penetrates the BBB because of protein binding⁴⁵, probably blocks the early phase of fever by acting on the luminal side of the BBB, thus decreasing secretion of prostaglandin E2 from the abluminal side⁴⁶.

Adaptations and modifications of BBB functions. The functions of the BBB are not static; rather, they adapt to the needs of the CNS throughout life and in response to disease. Both blood-to-brain and brain-to-blood transporters can rapidly adapt to physiological changes

or be affected by (and contribute to) disease states. Transporters seem to change much earlier, more rapidly, and in response to a larger number of conditions than do other features of the BBB, such as barrier integrity or transendothelial diffusion. Some of the most dramatic examples of BBB adaptation to the needs of the CNS occur during brain development and maturation⁴⁷. Such adaptations in transport activity are physiologically necessary so that the BBB can respond to the changing needs of the CNS. Altered transport can also be the result of disease, a cause of disease, or a sustaining element of disease.

BBB adaptations can be key elements in an organism's responses to physiological challenges. The rate at which the anorectic leptin is transported across the BBB⁴⁸, for example, is greatly reduced during starvation, while transport of the orexigenic peptide ghrelin is increased⁴⁹. By increasing the signals that promote feeding and decreasing those that suppress feeding, the BBB is acting to encourage changes in behaviour that are appropriate to starvation. Furthermore, the effects on the ghrelin and leptin transporters are mediated in part by serum triglycerides. Triglyceride levels are increased in blood as they are mobilized from fat as the main energy source in starvation. By inducing changes in the BBB transport rate of feeding hormones, hypertriglyceridemia is signalling the state of starvation to the brain.

BBB alterations are also seen after insults to the CNS. Spinal cord injuries alter the transport rates of molecules with neuroinflammatory and neurotrophic properties, such as TNF, PACAP and leukaemia inhibitory factor⁵⁰, whereas stroke upregulates transport of TNF⁵¹ and PACAP⁵². Inflammation increases the rate at which immune cells and pathogens such as HIV-1 can cross the BBB, and it can modulate many transport systems, including P-gp⁵³.

Some of these adaptations by the BBB are mediated by molecular signals coming from the CNS or peripheral tissues via the blood. Many molecules have been identified that can modulate transporter function and therefore could be the agents by which physiological states and diseases alter transporters. The effects of triglycerides released from adipose tissue on the transport of feeding hormones have been discussed above. Other examples include the ability of arginine vasopressin, insulin and amylin to alter the transport of amino acids across the BBB, and the ability of endothelin-1 to decrease — and of TNF to increase — P-gp activity at the BBB.

Four mechanisms have been elucidated by which molecules can alter BBB transporters. The first mechanism involves the molecule acting as an allosteric regulator of the transporter, as exemplified by the actions of aluminium and leucine on the BBB efflux transporter peptide transport system 1 (PTS-1). The second mechanism involves the activation of cellular mechanisms that alter the expression of the transporter, as exemplified by LPS, endothelin-1 and TNF altering P-gp function through nuclear factor- κ B (NF- κ B)^{54,55}. The third mechanism involves stimulating the release of another substance, as exemplified by LPS enhancing HIV-1

transport through the release of IL-6 and GM-CSF³⁷. The fourth mechanism is competitive inhibition, and is exemplified by verapamil increasing brain uptake of peptides, colchicine, and vinblastine; in this case, verapamil competes with these substances for brain-to-blood efflux by P-gp^{56,57}.

The adaptive mechanisms and their manipulation provide opportunities for drug development. Examples of therapeutic manipulation of transporters include increasing the delivery of leptin and lysosomal enzymes to the brain with adrenergic receptor modulators. Defective transport of leptin across the BBB is an early form of leptin resistance in obesity. α -adrenergic receptor modulators can increase the rate of leptin transport by two- to threefold, and this may be one mechanism by which drugs with activity at α -adrenergic receptors are associated with weight loss⁵⁸. The CNS manifestations of lysosomal storage diseases respond to enzyme replacement therapy during the neonatal period⁵⁹ because M6PR is active, but not during adulthood when M6PR activity at the BBB is abated. Treatment with adrenaline restores M6PR activity levels to those of the neonate and offers a strategy for treating these diseases in adults⁶⁰. Drugs that block P-gp (such as verapamil)³³ or alter its expression⁶¹ have been proposed as therapeutic approaches; the use of anti-seizure drugs that are not P-gp ligands⁶², such as valproic acid, might also provide a therapeutic strategy. Manipulation of serum triglyceride levels could shift the BBB transport of feeding hormones away from the starvation pattern; one study found that gemfibrozil, a drug that selectively lowers triglycerides, induces weight loss⁶³. Efflux transporters can also be manipulated, allowing increased entry of drugs that would otherwise be excluded. For example, the activities of P-gp and LRP1 can be modulated by endothelin 1, TNF, vitamin D and drugs that increase nitric oxide levels⁶⁴.

The ability to adapt provides a mechanism by which the BBB can fine-tune its service to the CNS. However, adaptive mechanisms can also be misdirected, leading to or supporting disease states. For example, triglycerides are elevated in starvation, but they are also elevated in obesity. Whereas triglycerides have an adaptive effect on the transport of feeding hormones in starvation, they have a maladaptive effect in obesity and may underpin the inability of peripherally administered hormones to effectively treat obesity⁶⁵. Seizures can increase P-gp activity and, because almost every anti-seizure drug is a weak P-gp ligand, they can reduce the ability of these drugs to reach the brain⁶⁶. Increased P-gp activity may explain why patients with status epilepticus and Lennox–Gastaut syndrome are refractory to anti-seizure medications^{66,67}. Alterations in transporter activities caused by diseases complicate drug delivery and development, suggesting that, when possible, drug development is best carried out using animal models of the disease of interest.

Classical CNS drug delivery revisited

The work reviewed above is influencing the more established approaches to drug development. The best understood mechanisms by which substances cross the

BBB are passive diffusion (also known as transendothelial diffusion) and the saturable processes of facilitated diffusion and active transport. Extracellular pathways and adsorptive transcytosis, mechanisms of entry that were not originally well differentiated from BBB disruption, are increasingly being investigated. Although many factors determine the degree to which a substance crosses the BBB, these factors largely fall into nine categories⁶⁸, as outlined below. Interactions among these nine mechanisms, especially in light of the new information reviewed above, make the development of CNS drugs complicated, but also offer interesting and unexpected opportunities (TABLE 1).

Passive diffusion. Passive diffusion is the non-saturable mechanism by which most small molecules in the pharmacopoeia enter the brain. Unlike uptake by most cells, which requires penetration of only the cell membrane, passage across the BBB requires that the entire cell be traversed. The degree to which a substance can undergo such transendothelial diffusion is largely dictated by its pharmacokinetic characteristics. The most important of these physicochemical characteristics are lipid solubility, hydrogen bonding and molecular mass^{69,70}. For a substance to reach its site of action in the CNS, it must partition from the lipid environment of the cell membranes that form the BBB back into the aqueous environment of brain interstitial fluid and/or CSF. Drugs that are extremely lipid soluble become sequestered in the peripheral and barrier cell membranes, reducing their ability to access brain interstitial fluid. As measured by the octanol/water partition coefficient, the ideal ratio for brain extraction is between 10:1 and 100:1. Charge influences partitioning between aqueous and lipid environments; because CSF tends to be slightly more acidic than plasma, uptake of weak bases is slightly favoured and weak acids are somewhat more excluded from CSF⁷¹. Charge also influences some of the other eight categories that determine the extent to which a drug can cross the BBB, especially those related to transporters, protein binding, and adsorptive transcytosis.

The relationships between BBB permeability and physicochemical characteristics can be predicted with greatest accuracy for small molecules, but similar — albeit modified — rules exist at least for small biologics, such as peptides^{72,73}. For example, BBB permeability will be influenced by secondary structure — for instance, folding to ‘hide’ charges, thereby making the peptide more lipophilic than the primary structure would predict⁷⁴. Cyclization greatly increases the lipid solubility of peptides, but often makes them substrates for brain-to-blood efflux systems so that net uptake by the brain is often decreased⁷⁵. Halogenation, the addition of lipid peptide or non-peptide co-drugs and signal peptides, and substitution of amino acids with less charged amino acids have all been shown to increase the BBB penetration of peptides. Despite the powerful influences of physicochemical characteristics on BBB penetration, the influence of the other eight categories limits the exclusive use of theoretical or *in silico* models in predicting drug action.

Passive diffusion

The mechanism by which a substance crosses the blood–brain barrier (BBB) by non-saturable means, with the degree of passage depending on the physicochemical characteristics of the substance.

Extracellular pathways

Areas such as the pial surface and subarachnoid space that are deficient in a blood–brain barrier (BBB) and thus allow small amounts of blood-borne substances, including albumin and immunoglobulins, to access the brain primarily through the Virchow–Robin spaces.

Cerebral blood flow. Substances can be broadly categorized as those that are flow dependent and those that are flow independent. Flow dependence occurs when a substance is extracted from the circulation so rapidly (either because of transendothelial diffusion or because of robust saturable transport) that the concentration at the venous end of the capillary is significantly lower than at the arterial end. For these substances, increasing blood flow increases the amount of drug presented to the BBB and thereby the amount of drug entering the brain. Glucose is flow dependent and so a brain region has to meet increased metabolic demands by increasing blood flow. Flow dependence is a theoretical disadvantage for drug delivery as blood flow, and thus the amount of drug delivered to a brain region, will depend on the metabolic activity of the brain region rather than on its therapeutic need. By contrast, the delivery rate of a flow-independent drug is the same regardless of the metabolic state of viable brain tissue. Such mismatches between therapeutic levels and blood flow would be particularly acute after ischaemic events and may underlie some of the toxicity associated with tissue-type plasminogen activator (tPA) in the treatment of stroke, as discussed below.

Degradation. The peripheral tissues, blood, CNS and the cells comprising the BBB can all be significant sources of drug degradation. Degradation by peripheral tissues and blood affect the pharmacokinetics of clearance before the drug can penetrate the BBB, and degradation within the CNS limits residence time at the site of action. Degradation at the barrier tissues can create an enzymatic barrier that can be a substantial reason for a substance not being taken up by the CNS, as exemplified by the monoamines and encephalins^{76,77}. The BBB acts as an enzymatic barrier to the monoamines adrenaline, noradrenaline and dopamine, in both the blood-to-brain and brain-to-blood directions, using the enzymes monoamine oxidase and catechol-*O*-methyltransferase⁷⁷. Drug penetration of the BBB and neurotoxicity can be influenced by these enzymes as illustrated by the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridines: the species differences in the ability of this drug to cause Parkinson disease seems to be caused by differences in the expression of monoamine oxidase⁷⁸. In perhaps the earliest example of the existence of the NVU, these enzymes are induced at the BBB by the adjacent neural tissues⁷⁹. In a similar fashion, the enzyme 4-aminobutyrate aminotransferase creates a barrier to GABA (γ -aminobutyric acid) permeation⁸⁰. In the case of encephalins, inhibition of enzymes or the use of enzymatically resistant analogues indicates that uptake is reduced to about one-quarter by the enzymatic barrier⁷⁶.

Pharmacokinetics of the substance. A substance cannot cross the BBB unless it can be presented to the BBB, and that presentation is largely dictated by the pharmacokinetic characteristics of the substance. Regardless of their peripheral route of administration (intradermal, intraperitoneal, intramuscular or intravenous), most drugs must enter the blood before they can enter the brain (other routes, such as retrograde nerve transmission, account for little of the uptake of most therapeutics). A substance with

a large volume of distribution or a short half-life in blood will have a lower concentration and less residence time in blood, thus decreasing its presentation to the luminal surface of the vascular BBB, than a substance with the opposite characteristics. Many biologics, including peptides and regulatory proteins, have very short half-lives, creating a pharmacokinetic 'barrier' to CNS penetration. Extending the half-life of a substance, for example, can proportionately increase BBB penetration by increasing its residence time in blood and thus prolonging its presentation to the BBB^{81,82}.

Binding in blood. As a general rule, only unbound or free material is available for crossing the BBB via transendothelial diffusion. Transporters are able to access bound materials if the affinity of the substance for the transporter greatly exceeds that of the binder. Many substances are weakly bound to proteins such as albumin, and this fraction of the substance is usually available to transporters, but not for passive diffusion. Thus, the 'biologically available' fraction includes both unbound material and material that is 'weakly' bound, with 'weakly' defined relative to the affinity of the transporter. Plasma and serum proteins constitute the traditional binders in blood, but some evidence shows that cells can also act as binders^{83,84}.

Protein binding, however, can have the paradoxical effect of increasing BBB penetration. This can occur because protein binding can greatly alter the peripheral pharmacokinetics of a drug, increasing its half-life and reducing its volume of distribution. Thus, as recently illustrated for IL-2 (REF. 85), protein binding can have a complex effect on brain uptake, slowing the rate of BBB penetration but increasing the time and concentration of drug exposure to the BBB. For compounds with extremely long half-lives, the extracellular pathways can become relevant for drug delivery⁸⁶.

Extracellular pathways. The extracellular pathways, or functional leaks, are a group of anatomically defined areas that enable small amounts of circulating substances to access the CNS. They include the large vessels of the pial surface and subarachnoid space, some of the interfaces of certain circumventricular organs, the sensory ganglia of spinal and cranial nerves, and the nucleus tractus solitarius⁸⁷. They are a route by which substances can reach the Virchow–Robin spaces, and hence participate in the glymphatic pathway, and are thought to be a major contributor to CNS levels of albumin.

The amounts of material entering the brain via the extracellular pathways are small, as exemplified by albumin with its CSF/serum ratio of about 0.005. However, it has been suggested that substances with long circulating half-lives, small volumes of distribution (that is, largely confined to the vascular space) and potent CNS effects can use the extracellular pathways as a therapeutic route⁸⁶. Antibodies, erythropoietin, and modified lysosomal enzymes are thought to enter the BBB-intact CNS by these routes.

Antibodies with the potential to cross the BBB are increasingly being developed as therapeutics. The mechanisms by which antibodies or antibody analogues cross

Table 1 | Strategies for improved delivery of therapeutics to the brain*

Strategy	Rationale	Challenges	Examples
Improve peripheral pharmacokinetics	Prolonged presentation to BBB increases opportunity to cross the BBB; very long circulation times in blood can allow the use of extracellular pathways	Complex, difficult to predict <i>in silico</i>	Lysosomal enzymes ⁸² ; leptin analogues ⁸³ ; G-CSF pegylation ¹¹³ ; antibodies ⁸⁶
Improve lipophilicity and/or hydrogen bonding of small peptides	Passive diffusion and transendothelial diffusion can occur for peptides <500 Da	Penetration rates are low, so need potent peptide analogues; increases uptake by peripheral tissues and so net effect on brain uptake may not be improved; structural changes can reduce affinity for CNS receptor; can become substrate for BBB efflux systems	Modify secondary structure to 'hide' charges ⁷⁴ ; decrease hydrogen bonding ⁷³
Use mechanism of adsorptive transcytosis	Glycoprotein–glycoprotein interactions can induce vesicular BBB transport; can also induce with highly charged small molecules (for example, polylysines) and other molecules (pluronics; fatty acids)	Difficult to predict or control intracellular vesicular trafficking; some glycoproteins and small highly charged molecules can be very toxic	Enhanced delivery with pluronics and glycosylation ^{114,117} ; delivery of lysosomal enzymes with charged molecules (takes advantage of the tendency of adsorptive transcytosis to route to lysosomes) ¹¹⁶ ; adaptation of mechanisms used by CNS-invading pathogens ²⁶
Develop non-flow-dependent drugs	Drug delivery is independent of variations in cerebral blood flow	Often a trade-off of lower uptake by the brain	Most biologics are not flow dependent
Target BBB transporters (influx)			
Develop analogues of the endogenous ligands of transporters	Transporter already targeted to desired brain region	Analogue must remain a ligand for a CNS receptor as well as a BBB transporter	APC ¹⁶²
Develop 'Trojan Horse' carrier molecules	Variety of approaches; could result in a universal carrier	Difficult to predict reaction of BECs (can induce adsorptive endocytosis pathways)	Antibody carrier designed using pharmacokinetic principles ^{27,28}
Induce antagonist or other therapeutic effect by blocking influx of endogenous ligand	Influx blocker not required to cross the BBB; blocker does not have to bind to or be active at CNS receptor	Little is known about BBB transporters or their binding sites, making rational development difficult	Block CNS uptake of toxic amino acids using other amino acids ¹⁹ ; block leptin entry into the brain to induce weight gain ²⁰
Initiate exploration for unknown transporters	Literature suggests that most BBB transporters are not yet discovered; many transporters were discovered serendipitously, with the proteins and genes unknown	Multi-disciplinary, coordinated effort is required; difficult to develop a universal strategy	Serendipitous discovery of many transporters, including for antisense oligonucleotides with unknown transporter protein, gene and endogenous ligand ^{10,13}
Target BBB transporters (efflux)			
Improve BBB drug penetration and CNS retention by inhibiting efflux of a drug with competitive inhibitors, small molecules, regulators or BBB-penetrating antisense oligonucleotides	Many potential drugs are efflux ligands and so they cannot accumulate in the brain	The efflux transporters involved for the drug must be known; must have the drug to inhibit efflux; often more than one efflux system is involved in the efflux of a drug; efflux transporters often have multiple ligands so that inhibition can affect many drugs and endogenous substances	Inhibition of PTS-6 with BBB-penetrating antisense oligonucleotides improves the therapeutic effects of PACAP27 (REF. 14); improved retention of anticonvulsants by competitively inhibiting their efflux ³³
BBB as a therapeutic target			
Protection from or reversal of BBB disruption	Many diseases involve BBB disruption	Disruption mechanisms are complex with few defined targets	APC protects BECs from tPA-induced apoptosis ¹⁶² ; protection of BBB in diabetes ¹⁰³
Modulation with small molecules and/or allosteric regulators	Rate of the transporter can be regulated by small molecules; usually does not require a small molecule to cross BBB; enhancers and inhibitors are available	Structures of most BBB transporters and their allosteric binding sites are unknown	Increased BBB penetration of anorectic leptin with α -adrenergic receptor agonist ⁵⁸ ; induction of transport of lysosomal enzymes across the BBB with an α -adrenergic receptor agonist ⁶⁰

Table 1 (cont.) | Strategies for improved delivery of therapeutics to the brain*

Strategy	Rationale	Challenges	Examples
BBB as a therapeutic target (cont.)			
Induction or inhibition of barrier cell secretions	Luminal targets can result in abluminal secretions, so the therapeutic agent does not have to cross the BBB	Most examples are from <i>in vitro</i> literature and may not translate to the <i>in vivo</i> setting	Inhibition of LPS-induced fever by blocking prostaglandin secretion from BECs ⁴³
Block glycoprotein binding sites for viruses	Possible use of small molecules (such as protamine sulfates) or known drugs (such as heparin); the blocker does not have to cross the BBB	The blocker will probably interact with numerous off-target binding sites	Blockade of HIV with mannose-6-phosphate ²⁴
Target cells of the neurovascular unit			
Target binding sites between the immune cell and BEC in multiple sclerosis	Blocks immune cell trafficking	Decreased availability of BECs compared with immune cells makes the development of BBB targets more cumbersome than immune cell targets	Natalizumab in multiple sclerosis ¹⁷⁴
Preserve pericyte function	Pericyte loss is implicated in BBB disruption of several diseases	Usually requires drugs that can cross the BBB	Metabolic carbonic anhydrase inhibitors in diabetes ¹¹¹
Use trafficking immune cells as drug delivery systems	Immune cells cross BBB via diapedesis to deliver payload of drugs to CNS; can reduce peripheral exposure to a drug	Targeting, drug loading and drug discharge within the CNS are challenging	Demonstrated to deliver neuroprotective nanozymes in a Parkinson disease model ¹¹⁵

APC, activated protein C; BBB, blood–brain barrier; BEC, brain endothelial cell; CNS, central nervous system; G-CSF, granulocyte colony-stimulating factor; LPS, lipopolysaccharide; PACAP27, pituitary adenylate cyclase-activating polypeptide 27; PTS-6, peptide transport system 6; tPA, tissue-type plasminogen activator. *This table emphasizes biologics.

the BBB is poorly investigated, but the extracellular pathways are one route by which they enter the CNS⁸⁷. Interestingly, IgG molecules are retained by the brain less effectively than albumin because of an efflux pump⁸⁸. IgM molecules enter the CNS by the same mechanism as IgG and albumin, but do not seem to be effluxed so their accumulation by the brain is higher than that of the smaller IgG molecules⁸⁹.

AE/AT. AE/AT are vesicle-based processes induced by the binding of glycoproteins or highly charged molecules to the cell membrane of the BEC^{90,91}. Vesicles are often routed to lysosomes and back to the cell surface of origin or to intracellular organelles⁹². However, in some cases, vesicles originating from the luminal cell membrane are routed to the abluminal membrane, thus producing a route for the passage of the vesicular contents across the BBB. Some viruses and other pathogens are able to use mechanisms related to AE/AT to gain entry into the BEC or to completely cross it, thus gaining entry into the CNS. An important feature of AE/AT that distinguishes it from classic saturable transporters is that, rather than increasing amounts of ligand leading to a decrease in the percentage of its uptake, an increasing amount of ligand can further enhance uptake⁹³. In extreme cases, AE/AT can result in disruption of the BBB, supposedly by the coalescing of vesicles to form channels or canaliculi across the BEC⁹⁴.

It is likely that AE/AT-like mechanisms underlie many of the strategies used to enhance drug delivery to the brain⁹⁵. Many of the penetrating peptides, such as derivatives of Tat, are highly charged and are therefore likely to be acting through this mechanism. Antibodies that target receptor-mediated transporters, especially those antibodies targeting a region that is not the binding site of the endogenous ligand, probably induce AE/AT

rather than receptor-mediated transcytosis. This would explain why routing to lysosomes often complicates this approach. Likewise, many of the vesicular nanoparticles, especially those studded with various ‘targeting’ moieties, are probably inducing AE/AT. Whereas antibodies cross the BBB through the extracellular pathways, various antibody fragments can cross the BBB at rates faster than albumin, apparently by using mechanisms akin to adsorptive transcytosis^{96,97}.

BBB disruption. Disruption of the BBB has been repeatedly attempted as a mechanism for drug delivery. Disruption with osmotic opening can produce effective results, as shown by its application to the treatment of CNS tumours⁹⁸. However, such disruption requires careful management and its success might rely to some degree on the blood–tumour barrier being more sensitive to disruption than the barrier of the surrounding normal brain tissue. Pharmacological agents and ultrasound are also being explored as ways to open the BBB⁹⁹. Approaches that disrupt an intact BBB in an attempt to let in a candidate drug also let in circulating substances that are normally excluded by the BBB and can be quite toxic to the CNS. Interestingly, a study showed that BBB disruption in animal models of multiple sclerosis and Alzheimer disease resulted in little or no increase in the brain uptake of small, lipid-soluble molecules¹⁰⁰. This could be true of therapeutic disruption strategies as well because even a disrupted BBB is still relatively tight in comparison with non-barrier capillary beds. For substances that are efflux ligands, the magnitude of the additional entry resulting from BBB disruption as a rule cannot overcome the much more robust efflux effect¹⁰¹. As such, attempts to disrupt the BBB in the service of drug delivery have inevitably produced disappointing results.

Neurovascular hypothesis

The hypothesis that the impaired ability of the blood–brain barrier (BBB) to remove amyloid- β peptide from the central nervous system (CNS) contributes to amyloid- β accumulation and the progression of Alzheimer disease.

BBB disruption is a serious consequence of many disease states, including stroke, diabetes, seizures, hypertensive encephalopathy, AIDS, and brain and spinal cord injuries, including traumatic brain injuries; BBB disruption is also observed in animal models of chronic pain and intense inflammation. Pericyte loss seems to be a prerequisite for BBB disruption in at least some of the chronic diseases^{102–104}. Conditions associated with increased immune cell trafficking into the CNS, such as AIDS and multiple sclerosis, have evidence of BBB disruption that seems to be caused by intense diapedesis^{105,106}. Most studies have focused on tight junction dysfunction as the cause of BBB disruption, but re-induction of macropinocytotic and cannalicular pathways occurs as well^{107,108} and may even be the main cause of disruption. Most therapeutic development is focused on repairing the consequences of BBB disruption, such as the reduction of oxidative stress in the CNS. However, some work suggests that BBB disruption itself may be reversible or preventable. For example, decreasing nitric oxide in thiamine deficiency¹⁰⁹, protecting pericytes with MMP inhibitors in Alzheimer disease^{6,110} or with mitochondrial carbonic anhydrase inhibitors in diabetes¹¹¹, and inhibiting cyclooxygenase in neuroinflammation¹¹² can prevent or reverse BBB disruption.

Molecular modifications to enhance BBB penetration.

Numerous modifications have been applied principally to biologics in the hope of enhancing their penetration across the BBB. These include pegylation, esterification, addition of fatty acids, insertion of D-amino acids, reversal of their primary amino acid sequence, production of nanoparticles, and glycosylation with glucose or other sugars. The rationale for these approaches is varied and includes increasing transmembrane diffusion, increasing enzymatic resistance, producing more favourable pharmacokinetics, or taking advantage of a transport system. This is a promising but difficult area of drug development, as modifications often have effects on BBB permeability by mechanisms that are different from those intended. For example, pegylation — the attachment of polyethylene glycol polymers — is often used to enhance the penetration of a biologic across the BBB, but it can also retard penetration^{20,113}. Furthermore, the mechanism by which an effect on BBB penetration is produced is not always the most obvious one. For example, addition of a glucose moiety to a peptide can greatly enhance the penetration of the peptide across the BBB, but this is not because the peptide is now able to use the BBB glucose transporter GLUT1 (REF. 114). Increasing lipid solubility of a substance can actually decrease its ability to reach its CNS target because of increased uptake by peripheral tissues, because it becomes a P-gp substrate, or because it becomes trapped in the lipid membranes of the cells that constitute the BBB. Thus, it can be difficult to explain in retrospect, much less predict, what effect a modification will have on brain uptake. These examples emphasize that the development of BBB-permeating drugs is currently an experimental science, rather than one fully based on theoretical principles.

As discussed above, AE/AT is a mechanism that probably underlies the effect of many of the various drug modifications (particularly those involving glycosylation or highly charged molecules). Diapedesis, the process by which immune cells cross the BBB, has many similarities to AE/AT; for instance, diapedesis involves interactions between glycoproteins on the surface of the immune cell and glycoproteins on the luminal surface of the BEC. These similarities may be exploited in the development of nanoparticles that can be delivered to the CNS as well as in the development of immune cells as drug delivery vehicles¹¹⁵. Similarly, a better understanding of how pathogens use AE/AT to enter the brain could provide new strategies for drug delivery²⁶. Highly charged molecules and some other types of molecules (for example, pluronics) are able to induce an AT-like penetration of the BBB and have been used to deliver lysosomal enzymes¹¹⁶ and other biologics to the brain¹¹⁷.

The diseased BBB: a therapeutic target

The BBB is increasingly viewed as a central player in several disease states. BBB disruption in conditions such as stroke and enhanced permeability to immune cell trafficking in multiple sclerosis are classic examples. More recently, transporter dysfunction such as impaired glucose transport in De Vivo disease¹¹⁸, impaired thyroid hormone transport in Allan–Herndon–Dudley syndrome, and enhanced P-gp function in epilepsy resistant to anti-seizure medications⁶² have been recognized. The discussion above has reviewed several other ways in which the BBB can be involved in diseases, thus illustrating that the BBB can itself be a therapeutic target. Targeting the BBB offers many opportunities (TABLE 2). One advantage is that the BBB does not have to be crossed if the therapeutic can act on its luminal surface. Strategies for targeting the BBB include using the regulators of transporters to modify their activity, inhibiting efflux systems, direct treatment of underlying BBB dysfunction when the BBB is a source of disease, and modulating BBB secretions. Here, I expand on the role of the BBB in disease and as a therapeutic target by considering four conditions: Alzheimer disease, multiple sclerosis, diabetes and obesity (FIG. 3).

Alzheimer disease. Several mechanisms have been proposed that relate BBB or cerebrovascular dysfunctions to Alzheimer disease, including decreased cerebral blood flow, impaired glucose transport, BBB disruption, increased capillary tortuosity with altered rheology, and neurotoxic secretions from BECs¹¹⁹. These mechanisms are not mutually exclusive and many of them would reinforce the pathological effects of the others. Whereas some mechanisms could sustain pathological processes once set in place, others could initiate events that may lead to Alzheimer disease. As such, the BBB may have a fundamental, even causal, role in Alzheimer disease. Here, I consider three mechanisms in some detail: decreased bulk flow, the neurovascular hypothesis and decreased insulin transport (FIG. 3a).

Decreased bulk flow ultimately occurs because of reduced production of CSF. CSF is primarily produced at the choroid plexus and reabsorbed through drainage

Table 2 | BBB-based strategies for specific disease targets

Strategy	Rationale	Refs
Diabetes (CNS manifestations)		
Target CNS control of blood glucose and/or lipolysis (for example, increase transport of insulin across the BBB)	Improve blood glucose and lipid homeostasis	148,149
Reduce oxidative stress in pericytes (for example, via mitochondrial carbonic anhydrase inhibition)	Reduce BBB disruption	103
Alzheimer disease and vascular dementias		
Stimulate CSF bulk flow and glymphatic circulation	Clear A β and other toxins from the CNS	122
Enhance LRP1 and P-gp activities at the BBB	Clear A β from the brain	12,126
Decrease oxidation of LRP1 and other CNS proteins	LRP1 oxidation is increased in Alzheimer disease; antioxidants improve cognition	13,130
Prevent pericyte loss	Pericyte loss results in BBB disruption	102,135
Increase transport of insulin or other gastrointestinal hormones via the BBB into the brain	Gastrointestinal hormones in the CNS improve cognition	138,140
Obesity		
Develop leptin agonists that cross the BBB independently of the leptin transporter	Bypasses leptin peripheral resistance secondary to the BBB	117
Develop enzymatically resistant leptin (or other anorectics)	Improved pharmacokinetics increases presentation to the BBB	81
Lower blood triglycerides	Enhances transport of anorectic leptin and inhibits transport of orexigenic ghrelin across the BBB	49
Enhance adrenergic tone	Enhances transport of anorectic leptin across the BBB	58
Anorexia		
Block leptin BBB transporter or receptor (for example, with leptin analogue)	Increases appetite by blocking the ability of leptin to enter or act on the brain	20
Lysosomal storage diseases		
Re-induce mannose-6-phosphate transporter	Transports enzyme across the BBB	60
Enhance circulating half-life in blood	Enzyme can enter the brain by extracellular pathways	82
Drug-resistant seizures		
Decrease P-gp activity with inhibitors (for example, verapamil) or regulators (for example, vitamin D)	Decreases brain-to-blood efflux of anticonvulsant medications	33,61,73
Multiple sclerosis		
Target ICAM, VCAM and selectins on the BBB	Inhibits immune cell diapedesis	174

A β , amyloid- β ; BBB, blood–brain barrier; CNS, central nervous system; CSF, cerebrospinal fluid; ICAM, intercellular adhesion molecule; LRP1, low-density lipoprotein receptor-related protein 1; P-gp, P-glycoprotein; VCAM, vascular adhesion molecule.

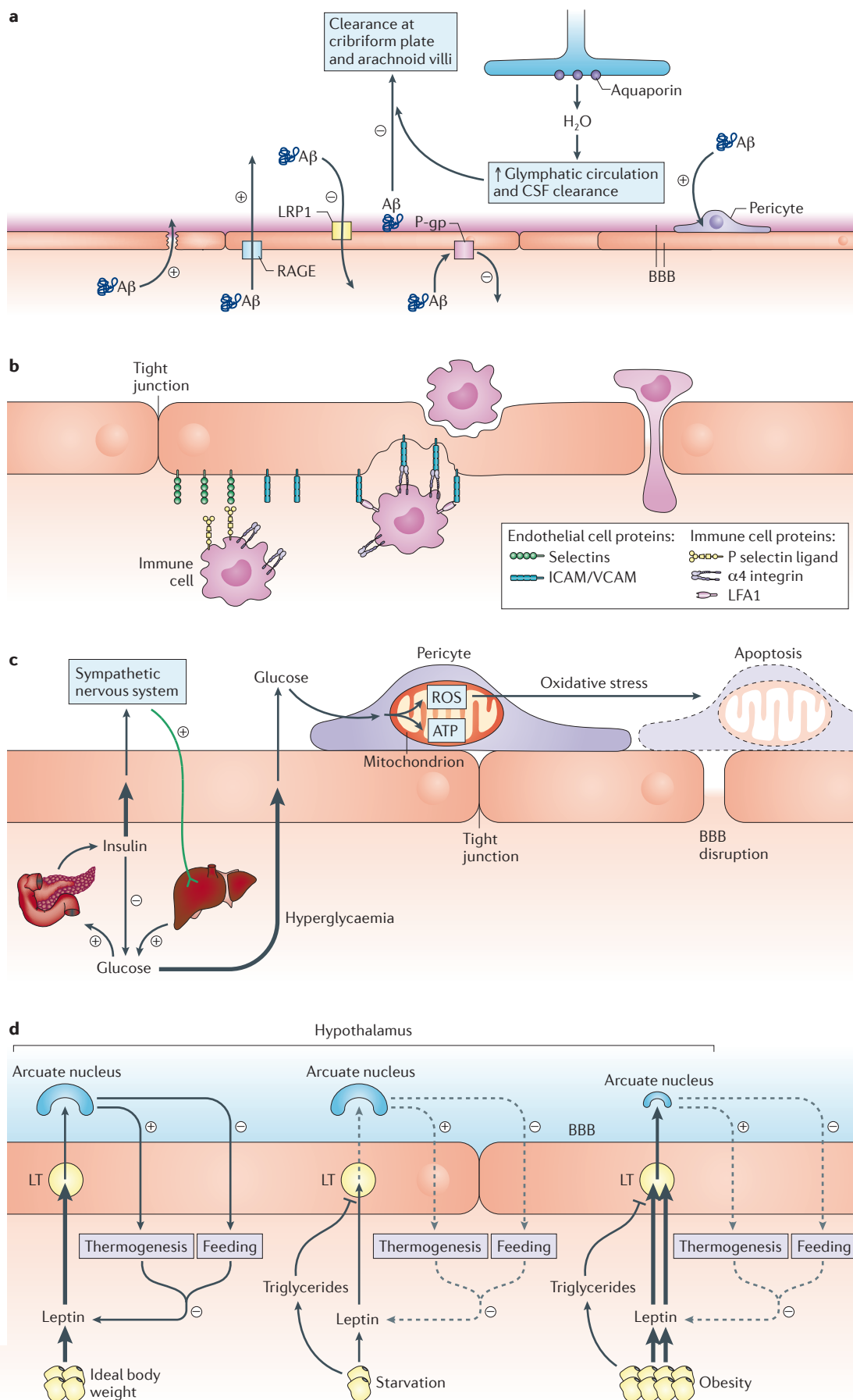
by the primitive lymphatics at the olfactory bulb and by the subarachnoid villi¹²⁰. With ageing and even more so in Alzheimer disease, the reabsorption of CSF back into the circulation is reduced¹²¹. With this slowing of reabsorption, CSF levels of albumin increase as do levels

of neurotoxic substances, including A β peptide. These processes can be further aggravated by a slowing of the glymphatic circulation¹²², a process that is regulated by astrocytic aquaporin, and by the BBB disruption that is characteristic of vascular dementia¹¹⁹. The trigger for decreased bulk flow in ageing and Alzheimer disease is not known, but inflammation can both decrease bulk flow and disrupt the BBB¹²³. Hence, therapeutic strategies suggested by this mechanism include blocking the neuroinflammation associated with Alzheimer disease¹²⁴, targeting astrocytic aquaporin production, and treating BBB disruption.

The neurovascular hypothesis states that decreased BBB clearance of A β leads to an increased amyloid burden in the brain and the progression of Alzheimer disease¹²⁵. The brain-to-blood efflux transporters P-gp and LRP1 are major regulators of CNS levels of A β ^{126,127} — a substance that, according to the amyloid hypothesis, has a direct, causal role in the onset and progression of Alzheimer disease. Patients with Alzheimer disease show decreased protein levels of LRP1 (REF. 128) and of P-gp¹²⁹ in their barrier cells, increased oxidation of LRP1 (REF. 130), and a decreased level of P-gp activity¹³¹. Knockdown of LRP1 in mice recapitulates the key features of Alzheimer disease predicted by the neurovascular hypothesis: decreased A β clearance from the brain, increased brain levels of A β , and impaired cognition¹³². Once initiated, decreased clearance of A β can become a self-perpetuating process as A β can further impair LRP1 function through oxidation¹³³. Decreasing A β levels with the peripheral administration of a phosphorothioate antisense oligonucleotide directed at amyloid precursor protein and capable of crossing the BBB restores A β clearance and cognition in animal models of Alzheimer disease¹³⁴.

Inflammation also impairs A β clearance from the brain and increases pericyte uptake of A β ¹²³. This inflammation-induced increase in uptake could be a mechanism accounting for the loss of pericytes seen in Alzheimer disease, given that A β is toxic to pericytes¹⁰². In animal models, the double hit of BBB dysfunction and pericyte loss results in a cycle of increasing pericyte loss, BBB damage and A β accumulation, leading to tau-related pathology, neuronal loss and cognitive impairments¹³⁵. This scenario shows how an impairment starting at the level of the BBB can promote the progression of or even initiate disease, as well as how the disease can lead to BBB impairment. It also provides several BBB-related therapeutic strategies, including enhancement of LRP1 and P-gp function and protection of pericytes from inflammation, oxidative stress and A β toxicity.

Insulin within the brain has very different roles compared with insulin in the periphery. In the brain, insulin has neurotrophic and neuroprotective functions. As the pancreas is considered essentially the sole source of insulin, the brain is dependent on BBB transport for its insulin. A decreased CSF/serum ratio suggests that the transport of insulin across the BBB is deficient in Alzheimer disease¹³⁶. Such a defect in transport would reinforce the insulin resistance recently shown to occur in the brains of patients with Alzheimer disease¹³⁷. Consistent with this CNS deficiency of insulin action, delivery of insulin



◀ **Figure 3 | The diseased blood–brain barrier.** **a** | Alzheimer disease. The blood–brain barrier (BBB) influences brain levels of amyloid- β (A β) peptide both by allowing its entry into the brain through receptor for advanced glycation end products (RAGE)-dependent transport and possibly BBB disruption, and by clearing it from the brain. The processes that are enhanced (+) or inhibited (–) in Alzheimer disease all tend towards an increased A β burden in the brain, which can induce pericyte death, leading to BBB disruption and possibly contributing to perivascular deposition of A β . **b** | Immune cell trafficking and multiple sclerosis. Diapedesis of immune cells across the vascular BBB is a multi-stage process that involves intercellular communication, resulting in upregulation and binding of immune and endothelial cell glycoprotein receptors. An invagination of the endothelial cell pairs with protrusions of the immune cell that eventually leads to the immune cell tunnelling through the endothelial cell. Immune cells can also cross the BBB through the paracellular route, which involves dissolution of brain endothelial cell tight junctions (as shown on the right). These processes are enhanced in multiple sclerosis, contributing to neuroinflammation. **c** | Diabetes. The pathway on the left shows that a negative feedback loop between insulin and glucose maintains euglycaemia. Insulin transported across the BBB influences sympathetic nervous system outflow, which controls liver glucose output. In diabetes, the negative feedback loop fails because of the loss of insulin (type 1 diabetes) and/or resistance to the actions of insulin (type 2 diabetes). Higher levels of glucose in the central nervous system (CNS) result in more ATP and reactive oxygen species (ROS). Pericytes are very sensitive to excess oxidative stress and their loss results in BBB disruption. **d** | Starvation, anorexia and obesity. The ideal body weight (left pathway) is maintained by a negative feedback loop in which leptin, secreted from adipose tissue, crosses the BBB via the leptin transporter (LT) and binds its receptor in the arcuate nucleus of the hypothalamus. CNS mechanisms shift behaviours away from feeding and towards calorically intense activities, reducing fat mass and leptin production. During starvation (middle pathway), leptin levels are reduced absolutely and per unit of fat. Triglycerides, which are elevated in the blood during starvation, interfere with leptin transport across the BBB. Feeding is increased and calorically expensive activities are suppressed (dashed arrows). Obesity (right pathway) is associated with elevations in blood leptin levels that are proportional to the increase in fat mass (double arrows), but the increase in cerebrospinal fluid (CSF) levels is proportionately lower because of both the saturable nature of the LT and transport inhibition by triglycerides. Resistance also develops at the leptin receptor at the arcuate nucleus. The brain stimulates feeding, resulting in an even greater fat mass. ICAM, intercellular adhesion molecule; LFA1, lymphocyte function-associated antigen 1; P-gp, P-glycoprotein; VCAM, vascular cell adhesion molecule.

to the CNS can improve some aspects of cognition in patients with Alzheimer disease within 15 minutes¹³⁸. The insulin transporter is a regulated system, with its rate being modulated by the NVU and circulating factors¹³⁹. Thus, insulin transport across the BBB could be increased by treatment with these modulators. Interestingly, almost every gastrointestinal peptide hormone has significant effects on cognition¹⁴⁰ and many of these hormones can cross the BBB. Analogues of these hormones that can cross the BBB may be useful not only in treating obesity and anorexia but also for treating dementias.

Multiple sclerosis. Multiple sclerosis is a neuroinflammatory disease characterized by demyelination and an intense CNS invasion of immune cells, especially lymphocytes and macrophages. The precipitating event or predisposing condition of multiple sclerosis is unknown, but penetration of the BBB by immune cells is both enhanced by and a cause of neuroinflammation, thus suggesting a mechanism by which multiple sclerosis can become self-perpetuating once the disease is established (FIG. 3b). BBB dysfunctions in experimental allergic encephalopathy, an animal model of multiple sclerosis, are characterized by increased diapedesis of immune cells, BBB disruption and altered cytokine transport^{141–143}.

Peripheral signals crossing the BBB may act on astrocytes and microglia, inducing them to secrete substances that then act on BECs to induce increased immune cell trafficking^{144,145}. Immune cell entry into the brain is key to sustaining the clinical symptoms of multiple sclerosis, as inhibition of such entry by blocking $\alpha 4$ integrin — the surface protein by which immune cells bind to vascular cell adhesion molecule 1 (VCAM1) on BECs — induces remission¹⁴⁶. However, immune cell invasion of the CNS involves a number of steps, each involving a variety of BBB receptors and transport mechanisms¹⁴⁷, and so provides multiple targets that could be used to modify immune cell trafficking into the CNS.

Diabetes. Two major types of BBB dysfunction have thus far been identified in diabetes. First, the CNS is increasingly recognized as a root cause of diabetes, especially type 2 diabetes, because of its involvement in controlling serum glucose levels and in its control of feeding behaviour and body weight (FIG. 3c). Altered transport of hormones can misinform the brain about nutrient status, thus fostering the development of hyperglycaemia and obesity. Second, diabetes results in BBB disruption.

The CNS is intimately involved in determining the serum level of glucose in many ways, including by controlling feeding behaviour, influencing glucose uptake by peripheral tissues, and regulating hepatic glucose production¹⁴⁸. The CNS monitors and integrates key peripheral signals, many of which must cross the BBB to convey their information to the CNS. As reviewed recently, many of these peripheral signals are gastrointestinal hormones¹⁴⁸. Notably, various brain nuclei in the hypothalamus regulate peripheral glucose production in response to insulin¹⁴⁹, which is derived from circulating insulin and requires transport by a saturable transporter at the BBB. Insulin transport, in turn, is influenced by metabolic conditions and substances, including starvation, obesity, inflammatory status, and triglycerides^{48,49,150}. As many of these conditions and substances also influence insulin sensitivity, they may be acting in part by affecting the access of insulin to the CNS. Thus, an approach for the treatment of diabetes is to increase the BBB transport of insulin and other hormones that regulate, through the CNS, the serum levels of glucose. This could be done by developing analogues of the hormones that can more readily cross the BBB or, as discussed elsewhere, by manipulating the metabolic conditions or substances that regulate the transporters.

Like the blood–retinal barrier¹⁵¹, the diabetic BBB becomes disrupted in humans and animal models^{152,153}. In both cases, glucotoxicity primarily at the level of the pericyte underlies the disruption^{103,104}. BBB pericytes can be protected by blocking glycolysis with mitochondrial carbonic anhydrase inhibitors^{111,154}. Such inhibition prevents the mitochondria from producing bicarbonate, which is a necessary ingredient for the metabolism of glucose through the Krebs cycle, ultimately reducing oxidative stress¹⁵⁵. Carbonic anhydrases, as exemplified by acetazolamide, have been used for decades for a number of conditions and are currently being tested for the treatment of diabetic retinopathy¹⁵⁶. Unfortunately, carbonic anhydrase inhibitors have off-target effects and those that are specific

for mitochondria do not exist. The development of drugs that can cross the BBB and target BBB mitochondrial carbonic anhydrases could be useful for the treatment of diabetic retinopathy and BBB disruption.

Obesity. Obesity is currently thought to occur when the brain underestimates the energy reserves of the body (FIG. 3d). Leptin, a hormone that is secreted from fat and crosses the BBB, is one of the chief hormones that informs the brain about those reserves. Obesity occurs when leptin is not secreted, when leptin resistance develops, or when there is a signalling defect downstream of the leptin receptor. Leptin resistance, which seems to be a near-universal feature in human obesity, occurs at the level of the CNS receptor (that is, central

resistance) and at the level of the BBB transporter (that is, peripheral resistance). The causes of peripheral resistance are at least twofold. First, saturation of the transporter begins at leptin levels that occur at a body weight considered by the Western world to be in the normal range; hence, as obesity progresses and serum leptin levels increase, less and less additional leptin is able to enter the brain to overcome the ever-increasing central resistance. Second, the leptin transporter is influenced by many substances, including triglycerides¹⁵⁷, which are elevated in obesity (as well as in starvation), as well as glucose and insulin¹⁵⁸, which are also usually elevated in obesity. As discussed above, the phenomenon of peripheral leptin resistance offers several approaches to the treatment of obesity and anorexia.

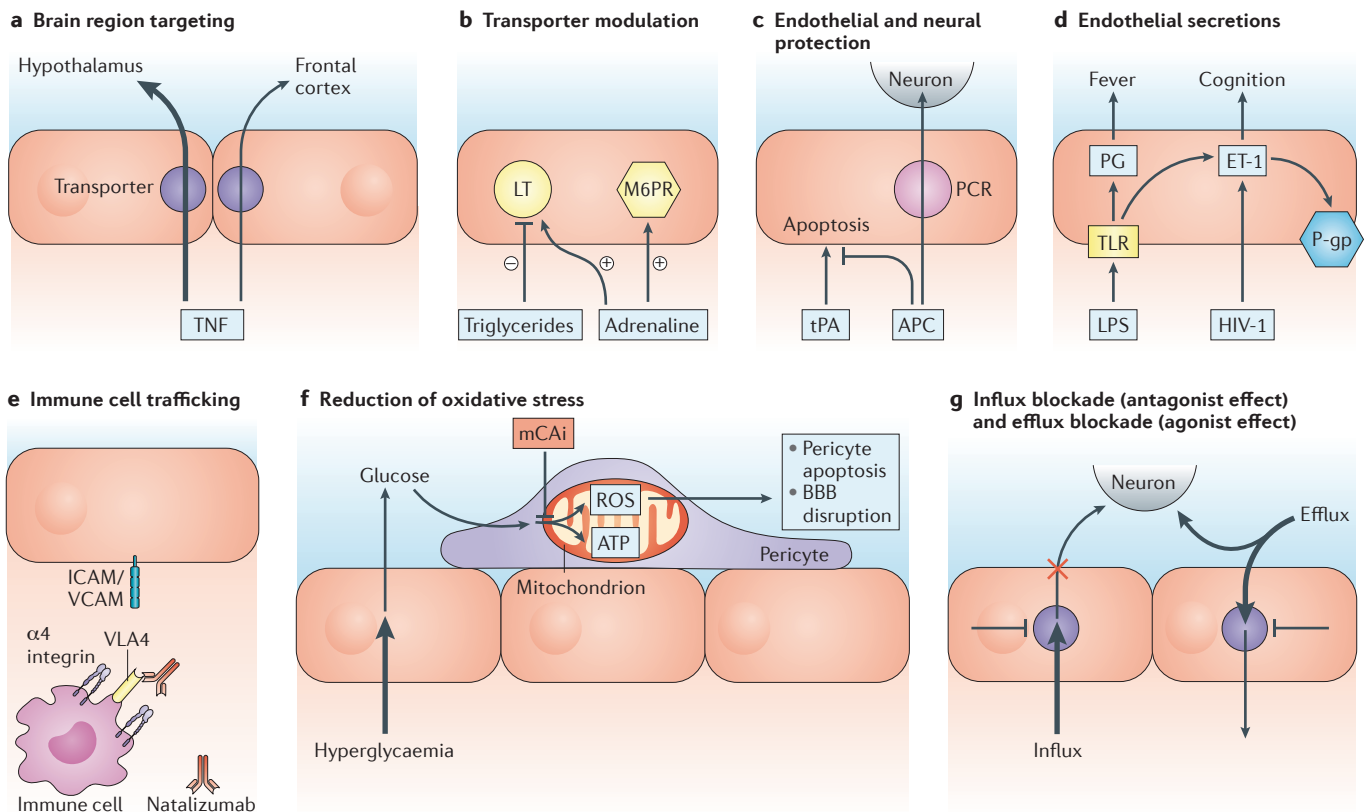


Figure 4 | Novel approaches to therapeutic targeting of the blood-brain barrier. **a** | Brain region targeting. Heterogeneous distributions of biologics transporters in the brain can be used to achieve targeted drug delivery. For example, the transport rate for tumour necrosis factor (TNF) is nine times greater into the hypothalamus than into the frontal cortex. **b** | Transporter modulation. Rates of transport can be enhanced or inhibited by modulators. For example, triglycerides inhibit and adrenaline stimulates leptin transporter (LT) activity, and adrenaline enhances the transport of lysosomal enzymes by the mannose-6-phosphate receptor (M6PR). **c** | Endothelial and neural protection. Many biologics have multiple actions. Activated protein C (APC) blocks tissue-type plasminogen activator (tPA) and crosses the blood-brain barrier (BBB) through the protein C receptor (PCR), thereby affording neuroprotection. **d** | Endothelial secretions. Stimulating or inhibiting abluminal secretions of blood endothelial cells (BECs) enables drugs to affect the central nervous system (CNS) environment without crossing the BBB. For example, inhibition of the binding of lipopolysaccharide (LPS) to Toll-like receptors (TLRs) or inhibition of prostaglandin (PG) synthesis can block

fever. P-glycoprotein (P-gp) activity is affected through several agents that converge on the production of endothelin-1 (ET-1) by BECs. Blockade of HIV-1 binding to the luminal surface of BECs or the inhibition of ET-1 release from the abluminal surface of BECs could reduce CNS levels of ET-1; ET-1 levels correlate with the severity of the cognitive impairment in AIDS. **e** | Immune cell trafficking. The multiple sclerosis drug natalizumab blocks immune cell trafficking into the CNS by binding to very late antigen 4 (VLA4), an $\alpha 4$ integrin, on immune cells. Intercellular adhesion molecule (ICAM) and vascular adhesion molecule (VCAM) receptors on the BBB offer additional therapeutic targets. **f** | Reduction of oxidative stress. High glucose levels in the brain result in increased electron transport chain (ETC) activity and production of reactive oxygen species (ROS), inducing pericyte apoptosis followed by BBB disruption. Mitochondrial carbonic anhydrase inhibitors (mCAi), by preventing bicarbonate production and thus the entry of pyruvate into the ETC, greatly decrease ROS. **g** | Blockade of influx removes agonist entry (producing a similar effect to delivering an antagonist into the brain) and inhibition of efflux results in an agonist effect by causing ligand retention in the brain.

Future directions: a combination of approaches

The areas of BBB research reviewed here show that the BBB is not simply a barrier that blocks drugs, but a complex, interactive, ever-adapting interface that serves the needs of the CNS and facilitates communication between the brain and peripheral tissues. This complexity offers an array of opportunities for drug development (FIG. 4). Below are other examples of unique approaches not considered above, including some that involve complex interactions between various functions of the BBB.

The direct involvement of the BBB in disease processes supports viewing the BBB as a therapeutic target. For example, downregulating the p75 neurotrophin receptor at the BBB improves outcomes in an animal model of multiple sclerosis¹¹. Vitamin D¹⁵⁹ and rosiglitazone¹⁶⁰ can increase LRP1 expression and efflux of A β , suggesting that these drugs may be able to reverse the neurovascular defects in Alzheimer disease¹²⁵. Estradiol reduces post-stroke oedema by downregulating the activity of a Na–K–Cl exchange channel at the BBB¹⁶¹.

Activated protein C (APC) illustrates how the BBB can have multiple therapeutic roles. tPA is used to dissolve the clots responsible for ischaemic stroke, but the drug can promote BEC apoptosis, increase the risk of brain haemorrhage, and is neurotoxic. APC protects BECs from post-ischaemic and tPA-induced apoptosis. Moreover, APC is transported across the BBB by the protein C receptor, where it protects neurons against tPA-induced and ischaemia-induced neurotoxicity. An APC analogue that retains the desired characteristics of BBB permeability, anti-apoptotic properties and neuroprotection has recently completed a Phase I safety trial¹⁶².

Control of BBB secretions, as discussed above, offers therapeutic opportunities, especially when combined with the concept of polarization. In one scenario, a drug acts at the luminal surface of the BBB to influence secretion from the abluminal surface directly into the brain. Thus, the drug is not required to cross the BBB and the

brain is, in effect, being treated by endogenous secretions. For example, amphotericin affects the release of prostaglandin from BECs¹⁶³ and *N*-acetylcysteine alters release of endothelin-1 from BECs¹⁶⁴, thus regulating P-gp activity⁵⁵. Many substances that affect P-gp function, including TNF and HIV-1, are known to release or even to act through endothelin-1 released from BECs^{55,165}. CSF endothelin-1 levels are elevated in HIV-1, ischaemic stroke and subarachnoid haemorrhage, and correlate with the severity of HIV encephalopathy¹⁶⁶. Thus *N*-acetylcysteine or other regulators of endothelin-1 release have the potential to affect a range of disease conditions while acting on the luminal surface of the BBB.

The NVU can also be involved in mediating the therapeutic effects of the BBB¹⁶⁷, as illustrated by the suppression of HIV replication in macrophages by BEC secretions. Activation of TLR3 on BECs induces the release of interferons, which, in turn, act on macrophages to suppress HIV-1 replication¹⁶⁸. Thus, HIV-1 replication in CNS macrophages could be attenuated by blocking TLR3 on the luminal surface of the BBB.

Conclusion

This Review has examined the main trends in research that have occurred over the past three decades involving the BBB as a regulatory interface and affecting drug delivery to the CNS. The BBB is intimately involved in crosstalk with the rest of the CNS and peripheral tissues and is crucial for normal brain functioning. Treating CNS diseases requires therapeutics that can cross or otherwise interact with the BBB. The complexity of the BBB, its regulatory mechanisms and its interactions with the other components of the NVU offer many new, unique opportunities for drug development. Additionally, the BBB can be a target of disease, and BBB dysfunctions can result in CNS diseases. As such, the BBB itself is a therapeutic target and is often more accessible to manipulation than the cells that it protects.

- Neuwelt, E. *et al.* Strategies to advance translational research into brain barriers. *Lancet Neurol.* **7**, 84–96 (2008).
- Hawkins, B. T. & Davis, T. P. The blood–brain barrier/neurovascular unit in health and disease. *Pharmacol. Rev.* **57**, 173–185 (2005).
- Daneman, R., Zhou, L., Kebede, A. A. & Barres, B. A. Pericytes are required for blood–brain barrier integrity during embryogenesis. *Nature* **468**, 562–566 (2010).
Demonstrates the role of pericytes in establishing the BBB and shows that barrier function is present even during the fetal period.
- Abbott, N. J., Ronnback, L. & Hansson, E. Astrocyte–endothelial interactions at the blood–brain barrier. *Nat. Rev. Neurosci.* **7**, 41–53 (2006).
- Banks, W. A. The blood–brain barrier in neuroimmunology: tales of separation and assimilation. *Brain Behav. Immun.* **44**, 1–8 (2015).
Review of mechanisms by which the BBB defines the neuroimmune system.
- Bell, R. D. *et al.* Apolipoprotein E controls cerebrovascular integrity via cyclophilin A. *Nature* **485**, 512–516 (2012).
- Oldendorf, W. H. Brain uptake of radio-labelled amino acids, amines and hexoses after arterial injection. *Am. J. Physiol.* **221**, 1629–1639 (1971).
- Kastin, A. J. & Pan, W. Blood–brain barrier and feeding: regulatory roles of saturable transport systems for ingestive peptides. *Curr. Pharm. Design* **14**, 1615–1619 (2008).
- Pan, W. & Kastin, A. J. Interactions of cytokines with blood–brain barrier: implications for feeding. *Curr. Pharm. Design* **9**, 827–831 (2003).
- Banks, W. A. *et al.* Delivery across the blood–brain barrier of antisense directed against amyloid β : reversal of learning and memory deficits in mice overexpressing amyloid precursor protein. *J. Pharmacol. Exp. Ther.* **297**, 1113–1121 (2001).
- Soilu-Hanninen, M. *et al.* Treatment of experimental autoimmune encephalomyelitis with antisense oligonucleotides against the low affinity neurotrophin receptor. *J. Neurosci. Res.* **59**, 712–721 (2000).
- Erickson, M. A. *et al.* Peripheral administration of antisense oligonucleotides targeting the amyloid- β protein precursor reverses A β PP and LRP-1 overexpression in aged SAMP8 mouse brain. *J. Alzheimers Dis.* **28**, 951–960 (2012).
- Farr, S. A. *et al.* Antisense oligonucleotide against GSK-3 β in brain of SAMP8 mice improves learning and memory and decreases oxidative stress: involvement of transcription factor Nrf2 and implications for Alzheimer's disease. *Free Radic. Biol. Med.* **67**, 387–395 (2013).
- Dogrukol-Ak, D. *et al.* Isolation of peptide transport system-6 from brain endothelial cells: therapeutic effects with antisense inhibition in Alzheimer's and stroke models. *J. Cereb. Blood Flow Metab.* **29**, 411–422 (2009).
- Poon, H. F. *et al.* Proteomic identification of less oxidized brain proteins in aged senescence-accelerated mice following administration of antisense oligonucleotide directed at the A β region of amyloid precursor protein. *Brain Res. Mol. Brain Res.* **138**, 8–13 (2005).
- Farr, S. A., Erickson, M. A., Niehoff, M. L., Banks, W. A. & Morley, J. E. Central and peripheral administration of antisense oligonucleotide targeting amyloid precursor protein improves learning and memory and reduces neuroinflammatory cytokines in Tg2576 (APPsw) mice. *J. Alzheimers Dis.* **40**, 1005–1016 (2014).
- Poduslo, J. F. & Curran, G. L. Permeability at the blood–brain barrier and blood–nerve barriers of the neurotrophic factors: NGF, CNTF, NT-3, BDNF. *Mol. Brain Res.* **36**, 280–286 (1996).
- Pan, W., Banks, W. A., Fasold, M. B., Bluth, J. & Kastin, A. J. Transport of brain-derived neurotrophic factor across the blood–brain barrier. *Neuropharmacology* **37**, 1553–1561 (1998).
- Fernstrom, J. D. Branched-chain amino acids and brain function. *J. Nutr.* **135**, 1439S–1546S (2005).
- Elinav, E. *et al.* Pegylated leptin antagonist is a potent orexigenic agent: preparation and mechanism of activity. *Endocrinology* **150**, 3083–3091 (2009).
- Georgieva, J. V., Hoekstra, D. & Zuhorn, I. S. Smuggling drugs into the brain: an overview of ligands targeting transcytosis for drug delivery across the blood–brain barrier. *Pharmaceutics* **6**, 557–583 (2014).

22. Pardridge, W. M. Blood–brain barrier drug delivery of IgG fusion proteins with a transferrin receptor monoclonal antibody. *Expert Opin. Drug Deliv.* **20**, 1–16 (2014).
23. Wang, T. et al. Toll-like receptor 3 mediates West Nile virus entry into the brain causing lethal encephalitis. *Nat. Med.* **10**, 1366–1373 (2004).
24. Dohgu, S., Ryser, J. S., Robinson, S. M. & Banks, W. A. Human immunodeficiency virus-1 uses the mannose-6-phosphate receptor to cross the blood–brain barrier. *PLoS ONE* **7**, e41623 (2012).
25. Hambleton, S. Chickenpox. *Curr. Opin. Infect. Dis.* **18**, 235–240 (2005).
26. McCall, R. L. et al. Pathogen-inspired drug delivery to the central nervous system. *Tissue Barriers* **8**, 3944449 (2014).
27. Atwal, J. K. et al. A therapeutic antibody targeting BACE1 inhibits amyloid- β production *in vivo*. *Sci. Transl. Med.* **3**, 84ra43 (2011).
28. Yu, Y. J. et al. Boosting brain uptake of the therapeutic antibody by reducing its affinity for a transcytosis target. *Sci. Transl. Med.* **3**, 84ra44 (2011).
29. Niewoehner, J. et al. Increased brain penetration and potency of a therapeutic antibody using a monovalent molecular shuttle. *Neuron* **81**, 49–60 (2014).
30. Moineddin, A., Morley, J. E. & Banks, W. A. Regional variations in the transport of interleukin-1 α across the blood–brain barrier in ICR and aging SAMP8 mice. *Neuroimmunomodulation* **8**, 165–170 (2000).
31. Maness, L. M., Banks, W. A., Zadina, J. E. & Kastin, A. J. Selective transport of blood-borne interleukin-1 α into the posterior division of the septum of the mouse brain. *Brain Res.* **700**, 83–88 (1995).
32. Begley, D. J. ABC transporters and the blood–brain barrier. *Curr. Pharm. Design* **10**, 1295–1312 (2004). **Classic and detailed description of P-gp and other ABC transporter CNS-to-blood efflux systems.**
33. Nicita, F. et al. Efficacy of verapamil as an adjunctive treatment in children with drug-resistant epilepsy: a pilot study. *Seizure* **23**, 36–40 (2014).
34. Faraci, F. M. & Heistad, D. D. Regulation of the cerebral circulation: role of endothelium and potassium channels. *Physiol. Rev.* **78**, 53–97 (1998).
35. Reyes, T. M., Fabry, Z. & Coe, C. L. Brain endothelial cell production of a neuroprotective cytokine, interleukin-6, in response to noxious stimuli. *Brain Res.* **851**, 215–220 (1999).
36. Kis, B. et al. Cerebral endothelial cells are a major source of adrenomedullin. *J. Neuroendocrinol.* **14**, 283–293 (2002).
37. Dohgu, S., Fleegal-DeMotta, M. A. & Banks, W. A. Lipopolysaccharide-enhanced transcellular transport of HIV-1 across the blood–brain barrier is mediated by luminal microvessel IL-6 and GM-CSF. *J. Neuroinflamm.* **8**, 167 (2011).
38. Cao, C., Matsumura, K., Yamagata, K. & Watanabe, Y. Involvement of cyclooxygenase-2 in LPS-induced fever and regulation of its mRNA by LPS in the rat brain. *Am. J. Physiol.* **272**, R1712–R1725 (1997).
39. Dohgu, S. & Banks, W. A. Brain pericytes increase the lipopolysaccharide-enhanced transcytosis of HIV-1 free virus across the *in vitro* blood–brain barrier: evidence for cytokine-mediated pericyte-endothelial cell cross talk. *Fluids Barriers CNS* **10**, 23 (2013).
40. Deli, M. A., Abraham, C. R., Kataoka, Y. & Niwa, M. Permeability studies on *in vitro* blood–brain barrier models: physiology, pathology, and pharmacology. *Cell. Mol. Neurobiol.* **25**, 59–127 (2005). **An authoritative and thorough review on the utility of, and approaches to, the gold standard *in vitro* monolayer model of the vascular BBB.**
41. Johanson, C. E. In *Neuromethods; The Neuronal Microenvironment* (eds Boulton, A. et al.) 33–104 (The Humana Press, 1988).
42. Verma, S., Nakaoka, R., Dohgu, S. & Banks, W. A. Release of cytokines by brain endothelial cells: a polarized response to lipopolysaccharide. *Brain Behav. Immun.* **20**, 449–455 (2006).
43. Engstrom, L. et al. Lipopolysaccharide-induced fever depends on prostaglandin E2 production specifically in brain endothelial cells. *Endocrinology* **153**, 4849–4861 (2012). **An elegant example of the BBB relay arm of the neuroimmune axis: blood-borne LPS binds to the luminal surface of the BEC, stimulating release of PGE2 from the abluminal surface into the CNS, thus inducing fever.**
44. Banks, W. A. & Robinson, S. M. Minimal penetration of lipopolysaccharide across the murine blood–brain barrier. *Brain Behav. Immun.* **24**, 102–109 (2010).
45. Parepally, J. M., Mandula, H. & Smith, O. R. Brain uptake of nonsteroidal anti-inflammatory drugs: ibuprofen, flurbiprofen, and indomethacin. *Pharm. Res.* **23**, 873–881 (2006).
46. Morimoto, A., Murakami, N., Nakamori, T. & Watanabe, T. Evidence for separate mechanisms of induction of biphasic fever inside and outside the blood–brain barrier. *J. Physiol.* **383**, 629–637 (1987).
47. Saunders, N. R., Daneman, R., Dziegielewska, K. M. & Liddelow, S. A. Transporters of the blood–brain and blood–CSF interfaces in development and in the adult. *Mol. Aspects Med.* **34**, 742–752 (2013).
48. Kastin, A. J. & Akerstrom, V. Fasting, but not adrenalectomy, reduces transport of leptin into the brain. *Peptides* **21**, 679–682 (2000).
49. Banks, W. A., Burney, B. O. & Robinson, S. M. Effects of triglycerides, obesity, and starvation on ghrelin transport across the blood–brain barrier. *Peptides* **29**, 2061–2065 (2008).
50. Pan, W., Cain, C., Yu, Y. & Kastin, A. J. Receptor-mediated transport of LIF across blood–spinal cord barrier is upregulated after spinal cord injury. *J. Neuroimmunol.* **174**, 119–125 (2006).
51. Pan, W. et al. Stroke upregulates TNF α transport across the blood–brain barrier. *Exp. Neurol.* **198**, 222–233 (2006).
52. Somogyvari-Vigh, A., Pan, W., Reglodi, D., Kastin, A. J. & Arimura, A. Effect of middle cerebral artery occlusion on the passage of pituitary adenylate cyclase activating polypeptide across the blood–brain barrier in the rat. *Regul. Pept.* **91**, 89–95 (2000).
53. Yu, C. et al. Neuroinflammation activates Mdr1b efflux transport through NF κ B: promoter analysis in BBB endothelia. *Cell Physiol. Biochem.* **22**, 745–756 (2008).
54. Yu, C., Pan, W., Tu, H., Waters, S. & Kastin, A. J. TNF activates MDR1 (P-glycoprotein) in cerebral microvascular endothelial cells. *Cell Physiol. Biochem.* **20**, 853–858 (2007).
55. Bauer, B., Hartz, A. M. S. & Miller, D. S. Tumor necrosis factor α and endothelin-1 increase P-glycoprotein expression and transport activity at the blood–brain barrier. *Mol. Pharmacol.* **71**, 667–675 (2007).
56. Chikale, E. G., Burton, P. S. & Borchardt, R. T. The effect of verapamil on the transport of peptides across the blood–brain barrier in rats: kinetic evidence for an apically polarized efflux mechanism. *J. Pharmacol. Exp. Ther.* **273**, 298–303 (1995).
57. Drion, N., Lemaire, M., Lefauconnier, J. M. & Scherrmann, J. M. Role of P-glycoprotein in the blood–brain transport of colchicine and vinblastine. *J. Neurochem.* **67**, 1688–1693 (1996).
58. Banks, W. A. Enhanced leptin transport across the blood–brain barrier by α 1-adrenergic agents. *Brain Res.* **899**, 209–217 (2001).
59. Vogler, C. et al. Enzyme replacement in murine mucopolysaccharidosis type VII: neuronal and glial response to β -glucuronidase requires early initiation of enzyme replacement therapy. *Pediatr. Res.* **45**, 838–844 (1999).
60. Urayama, A., Grubb, J. H., Banks, W. A. & Sly, W. S. Epinephrine enhances lysosomal enzyme delivery across the blood–brain barrier by up-regulation of the mannose 6-phosphate receptor. *Proc. Natl Acad. Sci. USA* **31**, 12873–12878 (2007).
61. van Vliet, E. A. et al. COX-2 inhibition controls P-glycoprotein expression and promotes brain delivery of phenytoin in chronic epileptic rats. *Neuropharmacology* **58**, 404–412 (2010).
62. Loscher, W. & Potschka, H. Role of multidrug transporters in pharmacoresistance to antiepileptic drugs. *J. Pharmacol. Exp. Ther.* **30**, 7–14 (2002).
63. Robins, S. J., Collins, D., McNamara, J. R. & Bloomfield, H. E. Body weight, plasma insulin, and coronary events with gemfibrozil in the Veterans Affairs High-Density Lipoprotein Intervention Trial (VA-HIT). *Atherosclerosis* **196**, 849–855 (2007).
64. Mandi, Y. et al. Nitric oxide production and MDR expression by human brain endothelial cells. *Anticancer Res.* **18**, 3049–3052 (1998).
65. Banks, W. A. Is obesity a disease of the blood–brain barrier? Physiological, pathological, and evolutionary considerations. *Curr. Pharm. Design* **9**, 801–809 (2003).
66. Liu, J. Y. et al. Neuropathology of the blood–brain barrier and pharmacoresistance in human epilepsy. *Brain* **135**, 3115–3133 (2012).
67. Kumar, A., Tripathi, D., Paliwal, V. K., Neyaz, Z. & Agarwal, V. Role of P-glycoprotein in refractoriness of seizures to antiepileptic drugs in Lennox–Gastaut syndrome. *J. Child Neurol.* **30**, 223–227 (2014).
68. Greig, N. H., Brossi, A., Pei, X.-F., Ingram, D. K. & Soncrant, T. T. In *New Concepts of a Blood–Brain Barrier* (eds Greenwood, J. et al.) 251–264 (Plenum Press, 1995). **A clear, concise review of the major factors that control drug entry into the CNS.**
69. Cornford, E. M., Braun, L. D., Oldendorf, W. H. & Hill, M. A. Comparison of lipid-mediated blood–brain-barrier penetrability in neonates and adults. *Am. J. Physiol.* **243**, C161–C168 (1982).
70. Oldendorf, W. H. Lipid solubility and drug penetration of the blood–brain barrier. *Proc. Soc. Exp. Biol. Med.* **147**, 813–816 (1974).
71. Rall, D. P., Stabenau, J. R. & Zubrod, C. G. Distribution of drugs between blood and cerebrospinal fluid: general methodology and effect of pH gradients. *J. Pharmacol. Exp. Ther.* **125**, 185–193 (1959).
72. Banks, W. A. & Kastin, A. J. Peptides and the blood–brain barrier: lipophilicity as a predictor of permeability. *Brain Res. Bull.* **15**, 287–292 (1985).
73. Chikale, E. G., Ng, K. Y., Burton, P. S. & Borchardt, R. T. Hydrogen bonding potential as a determinant of the *in vitro* and *in situ* blood–brain barrier permeability of peptides. *Pharm. Res.* **11**, 412–419 (1994).
74. Gray, R. A. et al. Delta-sleep-inducing peptide: solution conformational studies of a membrane-permeable peptide. *Biochemistry* **33**, 1323–1331 (1994).
75. Begley, D. J. Strategies for delivery of peptide drugs to the central nervous system: exploiting molecular structure. *J. Control. Release* **29**, 293–306 (1994).
76. Brownson, E. A., Abbruscato, T. J., Gillespie, T. J., Hruby, V. J. & Davis, T. P. Effect of peptidases at the blood brain barrier on the permeability of enkephalin. *J. Pharmacol. Exp. Ther.* **270**, 675–680 (1994).
77. Hardebo, J. E. & Owman, C. In *Pathophysiology of the Blood–Brain Barrier* (eds Johansson, B. B., Owman, C. & Widner, H.) 41–55 (Elsevier, 1990).
78. Kalara, R. N., Mitchell, M. J. & Harik, S. I. Correlation of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine neurotoxicity with blood–brain barrier monoamine oxidase activity. *Proc. Natl Acad. Sci. USA* **84**, 3521–3525 (1987). **A classic example of the enzymatic barrier and an early example of the NVU.**
79. Svendgaard, N. A., Bjorklund, A., Hardebo, J. & Stenevi, U. Axonal degeneration associated with a defective blood–brain barrier in cerebral implants. *Nature* **225**, 334–336 (1975).
80. van Gelder, N. M. In *Brain Barrier Systems* (eds Lajtha, A. & Ford, D. H.) 259–271 (Elsevier, 1968).
81. Novakovic, Z. M., Anderson, B. M. & Grasso, P. Myristic acid conjugation of [D-Leu-4]-OB3, a biologically active leptin-related synthetic peptide amide, significantly improves its pharmacokinetic profile and efficacy. *Peptides* **62**, 176–182 (2014).
82. Grubb, J. H. et al. Chemically modified β -glucuronidase crosses blood–brain barrier and clears neuronal storage in murine mucopolysaccharidosis VII. *Proc. Natl Acad. Sci. USA* **105**, 2616–2621 (2008).
83. Drewes, L. R., Conway, W. P. & Gilboe, D. D. Net amino acid transport between plasma and erythrocytes and perfused dog brain. *Am. J. Physiol.* **2**, E320–E325 (1977).
84. Jacquez, J. A. Red blood cell as glucose carrier: significance for placental and cerebral glucose transfer. *Am. J. Physiol.* **246**, R289–R298 (1984).
85. Patel, A. et al. Soluble interleukin-6 receptor induces motor stereotypes and co-localizes with Gp130 in regions linked to cortico–striato–thalamo-cortical circuits. *PLoS ONE* **7**, e1623 (2012).
86. Banks, W. A. Are the extracellular pathways a conduit for the delivery of therapeutics to the brain? *Curr. Pharm. Design* **10**, 1365–1370 (2004).
87. Broadwell, R. D. & Sofroniew, M. V. Serum proteins bypass the blood–brain barrier for extracellular entry to the central nervous system. *Exp. Neurol.* **120**, 245–263 (1993). **The first elucidation of the extracellular pathways to the CNS. Later studies showed that therapeutic antibodies and other drugs with similar pharmacokinetic properties can use this route to access the CNS.**
88. Garg, A. & Balthasar, J. P. Investigation of the influence of FcRn on the distribution of IgG to the brain. *AAPS J.* **11**, 553–557 (2009).
89. Banks, W. A. et al. Anti-amyloid beta protein antibody passage across the blood–brain barrier in the SAMP8 mouse model of Alzheimer’s disease: an age related selective uptake with reversal of learning impairment. *Exp. Neurol.* **206**, 248–256 (2007).

90. Mellman, I., Fuchs, R. & Helenius, A. Acidification of the endocytic and exocytic pathways. *Annu. Rev. Biochem.* **55**, 663–700 (1986).
91. Hardebo, J. E. & Kahrstrom, J. Endothelial negative surface charge areas and blood–brain barrier function. *Acta Physiol. Scand.* **125**, 495–499 (1985).
92. Villegas, J. C. & Broadwell, R. D. Transcytosis of protein through the mammalian cerebral epithelium and endothelium: II. Adsorptive transcytosis of WGA-HRP and the blood–brain and brain–blood barriers. *J. Neurocytol.* **22**, 67–80 (1993).
93. Banks, W. A., Kastin, A. J. & Akerstrom, V. HIV-1 protein gp120 crosses the blood–brain barrier: role of adsorptive endocytosis. *Life Sci.* **61**, L119–L125 (1997).
94. Vorbrodt, A. W., Dobrogowska, D. H., Ueno, M. & Lossinsky, A. S. Immunocytochemical studies of protamine-induced blood–brain barrier opening to endogenous albumin. *Acta Neuropathol.* **89**, 491–499 (1995).
95. Herve, F., Ghinea, N. & Scherrmann, J. M. CNS delivery via adsorptive transcytosis. *AAPS J.* **10**, 455–472 (2008).
- A clear, lucid review of potential of adsorptive transcytosis for CNS drug delivery.**
96. Chekhonin, V. P., Kabanov, A. V., Zhirkov, Y. A. & Morozov, G. V. Fatty acid acylated Fab-fragments of antibodies to neurospecific proteins as carriers for neuroleptic targeted delivery in brain. *FEBS Lett.* **287**, 149–152 (1991).
97. Peter, J. C. et al. A pharmacologically active monoclonal antibody against the human melanocortin-4 receptor: effectiveness after peripheral and central administration. *J. Pharmacol. Exp. Ther.* **333**, 478–490 (2010).
98. Kroll, R. A. & Neuwelt, E. A. Outwitting the blood–brain barrier for therapeutic purposes: osmotic opening and other means. *Neurosurgery* **42**, 1083–1099 (1998).
99. Saaber, D., Wollenhaupt, S., Baumann, K. & Reichl, S. Recent progress in tight junction modulation for improving bioavailability. *Expert Opin. Drug Deliv.* **9**, 347–381 (2014).
100. Cheng, Z. et al. Central nervous system penetration for small molecule therapeutic agents does not increase in multiple sclerosis- and Alzheimer's disease-related animal models despite reported blood–brain barrier disruption. *Drug Metab. Dispos.* **38**, 135–161 (2010).
101. Somjen, G. G., Segal, M. B. & Herreras, O. Osmotic-hypertensive opening of the blood–brain barrier in rats does not necessarily provide access for potassium to cerebral interstitial fluid. *Exp. Physiol.* **76**, 507–514 (1991).
102. Sengillo, J. D. et al. Deficiency in mural vascular cells coincides with blood–brain barrier disruption in Alzheimer's disease. *Brain Pathol.* **23**, 303–310 (2012).
103. Price, T. O., Eranki, V., Banks, W. A., Ercal, N. & Shah, G. N. Topiramate treatment protects blood–brain barrier pericytes from hyperglycemia-induced oxidative damage in diabetic mice. *Endocrinology* **153**, 362–372 (2012).
104. Hammes, H. P. et al. Pericytes and the pathogenesis of diabetic retinopathy. *Diabetes* **51**, 3107–3112 (2002).
105. Avison, M. J. et al. Inflammatory changes and breakdown of microvascular integrity in early human immunodeficiency virus dementia. *J. Neurovirol.* **10**, 223–232 (2004).
106. Boven, L. A., Middel, J., Verhoef, J., De Groot, C. J. & Nottet, H. S. Monocytes infiltration is highly associated with loss of tight junction protein zonula occludens in HIV-1 associated dementia. *Neuropathol. Appl. Neurobiol.* **26**, 356–362 (2000).
107. Lossinsky, A. S., Vorbrodt, A. W. & Wisniewski, H. M. Ultrastructural studies of vesicular and canalicular transport structures in the injured mammalian blood–brain barrier. *Acta Neuropathol.* **61**, 239–245 (1983).
108. Wahl, M., Unterberg, A., Baethmann, A. & Schilling, L. Mediators of blood–brain barrier dysfunction and formation of vasogenic brain edema. *J. Cereb. Blood Flow Metab.* **8**, 621–634 (1988).
109. Beauchesne, E., Desjardins, P., Hazell, A. S. & Butterworth, R. F. eNOS gene deletion restores blood–brain barrier integrity and attenuates neurodegeneration in the thiamine-deficient mouse brain. *J. Neurochem.* **111**, 425–459 (2009).
110. Halliday, M. R. et al. Relationship between cyclophilin A levels and matrix metalloproteinase 9 activity in cerebrospinal fluid of cognitively normal apolipoprotein E4 carriers and blood–brain barrier breakdown. *JAMA* **70**, 1198–1200 (2013).
111. Shah, G. N. et al. Pharmacological inhibition of mitochondrial carbonic anhydrases protects mouse cerebral pericytes from high glucose-induced oxidative stress and apoptosis. *J. Pharmacol. Exp. Ther.* **344**, 637–645 (2013).
112. Candelario-Jalil, E. et al. Cyclooxygenase inhibition limits blood–brain barrier disruption following intracerebral injection of tumor necrosis factor- α in the rat. *J. Pharmacol. Exp. Ther.* **323**, 488–498 (2007).
113. Frank, T. et al. Pegylated granulocyte colony-stimulating factor conveys long-term neuroprotection and improves functional outcome in a model of Parkinson's disease. *Brain* **135**, 1914–1925 (2012).
114. Polt, R., Dhanasekaran, M. & Keyari, C. M. Glycosylated neuropeptides: a new vista for neuropsychopharmacology. *Med. Res. Rev.* **25**, 557–585 (2005).
115. Batrakova, E. V., Gendelman, H. E. & Kabanov, A. V. Cell-mediated drug delivery. *Expert Opin. Drug Deliv.* **8**, 415–433 (2011).
116. Meng, Y. et al. Effective intravenous therapy for neurodegenerative disease with a therapeutic enzyme and a peptide that mediates delivery to the brain. *Mol. Ther.* **22**, 547–543 (2014).
117. Yi, X. & Kabanov, A. V. Brain delivery of proteins via their fatty acid and block copolymer modifications. *J. Drug Target.* **21**, 940–955 (2013).
118. De Vivo, D. C. et al. Defective glucose transport across the blood–brain barrier as a cause of persistent hypoglycorrhachia, seizures, and developmental delay. *N. Engl. J. Med.* **325**, 703–709 (1991).
119. Erickson, M. A. & Banks, W. A. Blood–brain barrier dysfunction as a cause and consequence of Alzheimer's disease. *J. Cereb. Blood Flow Metab.* **33**, 1500–1513 (2013).
120. Boulton, M. et al. Contribution of extracranial lymphatics and arachnoid villi to the clearance of a CSF tracer in the rat. *Am. J. Physiol.* **276**, R818–R823 (1999).
121. Alafuzoff, I., Adolfsson, R., Grundke-Iqbal, I. & Winblad, B. Blood–brain barrier in Alzheimer dementia and in non-demented elderly. *Acta Neuropathol.* **73**, 160–166 (1987).
122. Iliff, J. J. et al. A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β . *Sci. Transl. Med.* **4**, 147ra111 (2012).
- Key study demonstrating that the glymphatic pathway is important to CSF and brain interstitial fluid circulations, and clearance of toxins from the CNS.**
123. Erickson, M. A. et al. Lipopolysaccharide impairs amyloid β efflux from brain: altered vascular sequestration, cerebrospinal fluid reabsorption, peripheral clearance and transporter function at the blood–brain barrier. *J. Neuroinflamm.* **9**, 150 (2012).
124. Grammas, P. Neurovascular dysfunction, inflammation and endothelial activation: implications for the pathogenesis of Alzheimer's disease. *J. Neuroinflammation* **8**, 26 (2011).
125. Zlokovic, B. V. Neurovascular mechanisms of Alzheimer's neurodegeneration. *Trends Neurosci.* **28**, 202–208 (2005).
- Introduction of the neurovascular hypothesis, which states that impaired BBB clearance of A β peptide from the CNS is a fundamental contributor to Alzheimer disease.**
126. Zlokovic, B. V., Deane, R., Sagare, A. P., Bell, R. D. & Winkler, E. A. Low density lipoprotein receptor related protein-1: a serial clearance homeostatic mechanism controlling Alzheimer's amyloid β -peptide elimination from the brain. *J. Neurochem.* **115**, 1077–1089 (2010).
127. Hartz, A. M. S., Miller, D. S. & Bauer, B. Restoring blood–brain barrier P-glycoprotein reduces brain amyloid- β in a mouse model of Alzheimer's disease. *Mol. Pharmacol.* **77**, 715–723 (2010).
128. Donahue, J. E. et al. RAGE, LRP-1, and amyloid-beta protein in Alzheimer's disease. *Acta Neuropathol.* **112**, 405–415 (2006).
129. Wijesuriya, J. C., Bullock, J. Y., Faull, R. L. M., Hladky, S. B. & Barrand, M. A. ABC efflux transporters in brain vasculature of Alzheimer's subjects. *Brain Res.* **1358**, 228–238 (2010).
130. Owen, J. B. et al. Oxidative modification to LDL receptor-related protein 1 in hippocampus from subjects with Alzheimer's disease: implications for A β accumulation in AD brain. *Free Radic. Biol. Med.* **49**, 1798–1803 (2010).
131. van Assema, D. M. et al. Blood–brain barrier P-glycoprotein function in Alzheimer's disease. *Brain* **135**, 181–189 (2012).
132. Jaeger, L. B. et al. Testing the neurovascular hypothesis of Alzheimer's disease: LRP-1 antisense reduces blood–brain barrier clearance, increases brain levels of amyloid- β protein, and impairs cognition. *J. Alzheimers Dis.* **17**, 553–570 (2009).
- Provides experimental support for the neurovascular hypothesis first proposed in reference 125.**
133. Butterfield, D. A. & Boyd-Kimball, D. The critical role of methionine 35 in Alzheimer's amyloid β peptide (1–42)-induced oxidative stress and neurotoxicity. *Biochim. Biophys. Acta* **1703**, 149–156 (2005).
134. Banks, W. A. et al. Impairments in brain-to-blood transport of amyloid- β and reabsorption of cerebrospinal fluid in an animal model of Alzheimer's disease are reversed by antisense directed against amyloid- β protein precursor. *J. Alzheimers Dis.* **23**, 599–605 (2011).
135. Sagare, A. P. et al. Pericyte loss influences Alzheimer-like neurodegeneration in mice. *Nature Commun.* **4**, 2932 (2013).
136. Craft, S. et al. Cerebrospinal fluid and plasma insulin levels in Alzheimer's disease: relationship to severity of dementia and apolipoprotein E genotype. *Neurology* **50**, 164–168 (1998).
137. Talbot, K. et al. Demonstrated brain insulin resistance in Alzheimer's disease patients is associated with IGF-1 resistance, IRS-1 dysregulation, and cognitive decline. *J. Clin. Invest.* **122**, 1316–1338 (2012).
138. Reger, M. A. et al. Intranasal insulin administration dose-dependently modulates verbal memory and plasma amyloid- β in memory-impaired older adults. *J. Alzheimers Dis.* **13**, 323–331 (2008).
139. Urayama, A. & Banks, W. A. Starvation and triglycerides reverse the obesity-induced impairment of insulin transport at the blood–brain barrier. *Endocrinology* **149**, 3592–3597 (2008).
140. Berthoud, H. R. Interactions between 'cognitive' and 'metabolic' brain in the control of food intake. *Physiol. Behav.* **91**, 486–498 (2007).
141. Butter, C., Baker, D., O'Neill, J. K. & Turk, J. L. Mononuclear cell trafficking and plasma protein extravasation into the CNS during chronic relapsing experimental allergic encephalomyelitis in B10.A/H mice. *J. Neurol. Sci.* **104**, 9–12 (1991).
142. Hsueh, H., Pan, W., Wu, X. & Kastin, A. J. Cessation of blood-to-brain influx of interleukin-15 during development of EAE. *J. Cereb. Blood Flow Metab.* **29**, 1568–1578 (2009).
143. Juhler, M. et al. Blood–brain and blood–spinal cord barrier permeability during the course of experimental allergic encephalomyelitis in the rat. *Brain Res.* **302**, 347–355 (1984).
144. Mishra, P. K. et al. Loss of astrocytic leptin signaling worsens experimental autoimmune encephalomyelitis. *Brain Behav. Immun.* **34**, 98–107 (2013).
145. Hudson, L. C., Bragg, D. C., Tompkins, M. B. & Meeker, R. B. Astrocytes and microglia differentially regulate trafficking of lymphocyte subsets across brain endothelial cells. *Brain Res.* **1058**, 148–160 (2005).
146. Stuve, O. The effects of natalizumab on the innate and adaptive immune system in the central nervous system. *J. Neurol. Sci.* **274**, 39–41 (2008).
147. Correale, J. & Villa, A. The blood–brain barrier in multiple sclerosis: functional roles and therapeutic targeting. *Autoimmunity* **40**, 148–160 (2007).
148. Sandoval, D. A., Obici, S. & Seeley, R. J. Targeting the CNS to treat type 2 diabetes. *Nat. Rev. Drug Discov.* **8**, 386–398 (2009).
- Shows the fundamental role of the BBB in controlling blood glucose levels via its transport of insulin into the CNS.**
149. Scherer, T. et al. Brain insulin controls adipose tissue lipolysis and lipogenesis. *Cell Metab.* **13**, 183–194 (2011).
150. Banks, W. A., DiPalma, C. R. & Farrell, C. L. Impaired transport of leptin across the blood–brain barrier in obesity. *Peptides* **20**, 1341–1345 (1999).
151. Romeo, G., Liu, W. H., Asnaghi, V., Kern, T. S. & Lorenzi, M. Activation of nuclear factor- κ B induced by diabetes and high glucose regulates a proapoptotic program in retinal pericytes. *Diabetes* **51**, 2241–2248 (2002).
152. Huber, J. D., VanGilder, R. L. & Houser, K. A. Streptozotocin-induced diabetes progressively increases blood–brain barrier permeability in specific brain regions in rats. *Am. J. Physiol.* **291**, H2660–H2668 (2006).

153. Starr, J. M. *et al.* Increased blood–brain barrier permeability in type II diabetes demonstrated by gadolinium magnetic resonance imaging. *J. Neurol. Neurosurg. Psychiatry* **74**, 70–76 (2003).
References 152 and 153 establish that BBB disruption occurs in diabetes.
154. Shah, G. N., Morofuji, Y., Banks, W. A. & Price, T. O. High glucose-induced mitochondrial resistance and reactive oxygen species in mouse cerebral pericytes is reversed by pharmacological inhibition of mitochondrial carbonic anhydrase: implications for cerebral microvascular disease in diabetes. *Biochem. Biophys. Res. Commun.* **440**, 354–358 (2013).
Demonstrates that BBB disruption occurs because of oxidative stress arising from excess mitochondrial respiration.
155. Kowluru, R. A. Diabetic retinopathy: mitochondrial dysfunction and retinal capillary cell death. *Antioxid. Redox Signal.* **7**, 1581–1587 (2005).
156. Weiwei, Z. & Hu, R. Targeting carbonic anhydrase to treat diabetic retinopathy: emerging evidences and encouraging results. *Biochem. Biophys. Res. Commun.* **390**, 368–371 (2009).
157. Banks, W. A. *et al.* Triglycerides induce leptin resistance at the blood–brain barrier. *Diabetes* **53**, 1253–1260 (2004).
158. Kastin, A. J. & Akerstrom, V. Glucose and insulin increase the transport of leptin through the blood–brain barrier in normal mice but not in streptozotocin-diabetic mice. *Neuroendocrinology* **73**, 237–242 (2001).
159. Ito, S. *et al.* 1 α ,25-dihydroxyvitamin D₃ enhances cerebral clearance of human amyloid- β peptide(1–40) from mouse brain across the blood–brain barrier. *Fluids Barriers CNS* **8**, 20 (2011).
160. Moon, J. H. *et al.* The effect of rosiglitazone on LRP1 expression and amyloid β uptake in human brain microvascular endothelial cells: a possible role of a low-dose thiazolidinedione for dementia treatment. *Int. J. Neuropsychopharmacol.* **1**, 1–8 (2011).
161. O'Donnell, M. E., Lam, T. I., Tran, L. Q., Foroutan, S. & Anderson, S. E. Estradiol reduces activity of the blood–brain barrier Na-K-Cl cotransporter and decreases edema formation in permanent middle cerebral artery occlusion. *J. Cereb. Blood Flow Metab.* **26**, 1234–1249 (2006).
162. Lyden, P. *et al.* Phase 1 safety, tolerability and pharmacokinetics of 3K3A-APC in healthy adult volunteers. *Curr. Pharm. Design* **19**, 7479–7485 (2013).
163. McGuire, T. R. *et al.* Release of prostaglandin E-2 in bovine brain endothelial cells after exposure to three unique forms of the antifungal drug amphotericin-B: role of COX-2 in amphotericin-B induced fever. *Life Sci.* **72**, 2581–2590 (2003).
164. Sury, M. D. *et al.* Evidence that N-acetylcysteine inhibits TNF- α -induced cerebrovascular endothelin-1 upregulation via inhibition of mitogen- and stress-activated protein kinase. *Free Radic. Biol. Med.* **41**, 1372–1383 (2006).
165. Didier, N., Banks, W. A., Creminon, C., Dereuddre-Bosquet, N. & Mabondzo, A. HIV-1-induced production of endothelin-1 in an *in vitro* model of the human blood–brain barrier. *Neuroreport* **13**, 1179–1183 (2002).
166. Rolinski, B. *et al.* Endothelin-1 elevated in the cerebrospinal fluid of HIV-infected patients with encephalopathy. *Infection* **27**, 244–247 (1999).
167. Vangilder, R. L., Rosen, C. L., Barr, T. L. & Huber, J. D. Targeting the neurovascular unit for treatment of neurological disorders. *Pharmacol. Ther.* **130**, 239–247 (2011).
168. Li, J. *et al.* Immune activation of human brain microvascular endothelial cells inhibits HIV replication in macrophages. *Blood* **121**, 2934–2942 (2013).
169. Kubie, L. S. & Shults, G. M. Studies on the relationship of the chemical constituents of blood and cerebrospinal fluid. *J. Exp. Med.* **42**, 565–591 (1925).
170. Pincus, J. B. & Kramer, B. Comparative study of the concentration of various anions and cations in cerebrospinal fluid and serum. *J. Biol. Chem.* **57**, 463–470 (1923).
171. Davson, H. & Smith, H. V. Physiological aspects of the penetration of drugs into the cerebrospinal fluid. *Proc. R. Soc. Med.* **50**, 963–966 (1957).
172. Roth, L. J. & Barlow, C. F. Drugs in the brain. *Science* **134**, 22–31 (1961).
173. Reese, T. S. & Karnovsky, M. J. Fine structural localization of a blood–brain barrier to endogenous peroxidase. *J. Cell Biol.* **34**, 207–217 (1967).
A classic paper demonstrating the ultrastructural basis for the BBB: the presence of tight junctions and decreased transcytotic vesicles.
174. Coisne, C., Mao, W. & Engelhardt, B. Cutting edge: natalizumab blocks adhesion but not initial contact of human T cells to the blood–brain barrier *in vivo* in an animal model of multiple sclerosis. *J. Immunol.* **182**, 5909–5913 (2009).

Acknowledgements

The author is supported by the US Department of Veterans Affairs and a grant from the US National Institute on Aging (grant R01 AG046619).

Competing interests statement

The author declares no competing interests.

CORRIGENDUM

Biopharma deal-making in 2015: changing the pharma landscape

Amanda Micklus and Steven Muntner

Nature Reviews Drug Discovery **15**, 78–79 (2016)

The value of M&A deals in Q4 2015 was incorrectly stated as \$689 billion. The correct value is \$187 billion. The text and figure have been corrected accordingly.

ERRATUM

Pioneering apoptosis-targeted cancer drug poised for FDA approval

Asher Mullard

Nature Reviews Drug Discovery **15**, 147–149 (2016)

Martin Dyer's affiliation was incorrect. He is at the University of Leicester, UK. This has been corrected in the online version of the article.

ONLINE CORRESPONDENCE

Nature Reviews Drug Discovery publishes items of correspondence online. Such contributions are published at the discretion of the Editors and can be subject to peer review. Correspondence should be no longer than 500 words with up to 10 references and should represent a scholarly attempt to comment on a specific Review or Perspective article that has been published in the journal. To view correspondence, please go to our homepage at: <http://www.nature.com/nrdd> and follow the link from the current table of contents. To cite correspondence, please use its doi number.

The following correspondence has recently been published:

Open Lab as a source of hits and leads against tuberculosis, malaria and kinetoplastid diseases

Lluís Ballell, Mike Strange, Nicholas Cammack, Alan H. Fairlamb and Leszek Borysiewicz

doi:10.1038/nrd.2016.51

This correspondence relates to the article:

Hit and lead criteria in drug discovery for infectious diseases of the developing world

Kei Katsuno, Jeremy N. Burrows, Ken Duncan, Rob Hooft van Huijsduijnen, Takushi Kaneko, Kiyoshi Kita, Charles E. Mowbray, Dennis Schmatz, Peter Warner and B. T. Slingsby

Nat. Rev. Drug Discov. **14**, 751–758 (2015)